



```
CC -1- SIMILARITY: CONTAINS 2 RNASE III DOMAINS.
CC -1- CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift in position 775.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF199011; AAF80558.1; -
DR EMBL: AJ242976; CAB45133.1; -
DR EMBL: AK001121; BAA91511.1; ALT_INIT.
DR EMBL: AF116910; AAD29637.1; ALT_FRAME.
DR InterPro: IPR001159; DS_RBD.
DR InterPro: IPR000999; RNase_3.
DR Pfam: PF00035; dsrm.1.
DR Pfam: PF00636; Ribonuclease_3; 2.
DR SMART: SM00358; DSRM; 1.
DR SMART: SM00535; RIBOC; 2.
DR PROSITE: PS50137; DS_RBD; 1.
DR PROSITE: PS00517; RNASE_3_1; 2.
DR PROSITE: PS50142; RNASE_3_2; 2.
KW Ribosome biogenesis; Hydrolyase; Nuclease; Endonuclease; Repeat;
KW RNA-binding; Nuclear protein; Alternative splicing.
FT DOMAIN 1 212 PRO-RICH.
FT DOMAIN 219 316 ARG-RICH.
FT DOMAIN 876 1056 RNASE III 1.
FT DOMAIN 1107 1233 RNASE III 2.
FT DOMAIN 1260 1334 DRBM.
FT VARSPLIC 285 353 REPERERHRHRDNRSPLESYKKEKRSGRSYGLSVPE
PAGCGELIGELIKNTDSMAPIELVNH -> S (IN
ISOGORM 2)
FT CONFLICT 166 174 YQRPGRISH -> RERERTSLE (IN REF. 2).
FT CONFLICT 612 612 L -> P (IN REF. 2).
FT CONFLICT 1020 1020 R -> P (IN REF. 1).
FT CONFLICT 1230 1230 I -> T (IN REF. 1).
SQ SEQUENCE 1374 AA; 159315 MW; ED6FDEA09F3B8092 CRC64;

Query Match 99.6%; Score 7471; DB 1; Length 1374;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1370; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 421 YSSDPMDOYGVSTVGTSLRDLVDKFEELGSRQEKAAKAPWEPPTKTLDEDLSS 480
DB 421 YSSDPMDOYGVSTVGTSLRDLVDKFEELGSRQEKAAKAPWEPPTKTLDEDLSS 480
QY 481 SESECCSDSDSTSSSDSVFVYIAEIKRKAHPRLDELVDYNDPGMNGPLCKGSA 540
DB 481 SESECCSDSDSTSSSDSVFVYIAEIKRKAHPRLDELVDYNDPGMNGPLCKGSA 540
QY 541 KARTGRHSIYGEBAIKRCRPMTNAGRLFEYRTVSPPTNFDLPRTVLEYDDEHY 600
DB 541 KARTGRHSIYGEBAIKRCRPMTNAGRLFEYRTVSPPTNFDLPRTVLEYDDEHY 600
QY 601 FEESMFAPAALPTNIPCLKYIRNIDYTHFTEAMPENVCYKGLFSLFPROLLEY 660
DB 601 FEESMFAPAALPTNIPCLKYIRNIDYTHFTEAMPENVCYKGLFSLFPROLLEY 660
QY 661 DWNLKGPFLFEDSPCCRFHFMRFRVFLPDGKEVLSMHQILLYLRCSKALVPEEET 720
DB 661 DWNLKGPFLFEDSPCCRFHFMRFRVFLPDGKEVLSMHQILLYLRCSKALVPEEET 720
QY 721 NMLQWEELEWQYAEBCGMIVTNPSTKPSVRIIDLDREOFNDVITPEIIVHFGIRPA 780
DB 721 NMLQWEELEWQYAEBCGMIVTNPSTKPSVRIIDLDREOFNDVITPEIIVHFGIRPA 780
QY 781 QLSYAGDPQYQIKLMSYVYKLRHLANS PKYKQTDOKLAOREALOKIKQKTMREVTV 840
DB 781 QLSYAGDPQYQIKLMSYVYKLRHLANS PKYKQTDOKLAOREALOKIKQKTMREVTV 840
QY 841 ELSQSGQWKGIRSDVOQHMMMLPYLTNHTRYHQCMLNHDKLGYPFODRCILQMLMTP 900
DB 841 ELSQSGQWKGIRSDVOQHMMMLPYLTNHTRYHQCMLNHDKLGYPFODRCILQMLMTP 900
QY 901 SHHLNFGMNDHARNSLSNCGIRQPKYGDKRKVVHMMRRKGINTLINMSRLGDDPTPS 960
DB 901 SHHLNFGMNDHARNSLSNCGIRQPKYGDKRKVVHMMRRKGINTLINMSRLGDDPTPS 960
QY 961 RINHELELEFGDAVVEFLTSVHLTYLFPSEEGGLATYTAIVONHMLAKKLELDR 1020
DB 961 RINHELELEFGDAVVEFLTSVHLTYLFPSEEGGLATYTAIVONHMLAKKLELDR 1020
QY 1021 FMLYAHGPDLCRESDDLHNAANCFEALIGAVYLEGSLEEKOLFGRLEFNDPLREVMIN 1080
DB 1021 FMLYAHGPDLCRESDDLHNAANCFEALIGAVYLEGSLEEKOLFGRLEFNDPLREVMIN 1080
QY 1081 YPLHPILOLPENPDROLIETSPVLQKLTREEEAGVLETHVRLLAFTLRVGFNHLTL 1140
DB 1081 YPLHPILOLPENPDROLIETSPVLQKLTREEEAGVLETHVRLLAFTLRVGFNHLTL 1140
QY 1141 GHNRAMEFLGDSIMQLVATYELFIHPDNDHGHITLRSLSVNNRQAKVAEELGMOEYA 1200
DB 1141 GHNRAMEFLGDSIMQLVATYELFIHPDNDHGHITLRSLSVNNRQAKVAEELGMOEYA 1200
QY 1201 ITNDKTRPVGLRTKTLADLLESLIALYTDKIDLEVYHTPMNCFEPRLKEFTLNDMD 1260
DB 1201 ITNDKTRPVGLRTKTLADLLESLIALYTDKIDLEVYHTPMNCFEPRLKEFTLNDMD 1260
QY 1261 PKSQLQCCILTRTEGKEPDIPLYKTLQTVGSHARTYVAVYFKGERIGCGKPSIQQA 1320
DB 1261 PKSQLQCCILTRTEGKEPDIPLYKTLQTVGSHARTYVAVYFKGERIGCGKPSIQQA 1320
QY 1321 EMGAAMALEKYNPQMAHOKRFRIGKTYROELKEMKEREHOREDEPDEDIKK 1374
DB 1321 EMGAAMALEKYNPQMAHOKRFRIGKTYROELKEMKEREHOREDEPDEDIKK 1374

RESULT 2
RNC_CAEEL STANDARD; PRT; 1086 AA.
ID RNC_CAEEL
AC 001326; 001327; Q90908;
DT 01-NOV-1997 (Ref. 35, Created)
DT 01-MAR-2002 (Ref. 41, Last sequence update)
DT 01-MAR-2002 (Ref. 41, Last annotation update)
DE Ribonuclease III (EC 3.1.26.3) (RNase III).
```

```
GN F26E4.10 OR F26E4.13.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdida; Rhabditoidea;
OC Rhabdilidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA Lighting J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RM [2]
RA REVISIONS.
RR Lighting J.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RM [3]
RP SEQUENCE OF Z27-1086 FROM N.A.
RX MEDLINE-20179701; PubMed-10713462;
RT Filippov V., Solovyev V., Filippova M., Gill S.S.;
RU "A novel type of RNase III family proteins in eukaryotes.";
Gene 245:213-221(2000).
CC -I FUNCTION: Involved in pre-tRNA processing. Cleaves double-strand RNA and does not cleave single-strand RNA (By similarity).
CC -I CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-phosphonomoester.
CC -I SUBCELLULAR LOCATION: Nuclear (Probable).
CC -I SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRAINED RNA-BINDING) DOMAIN.
CC -I SIMILARITY: CONTAINS 2 RNASE III DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/or send an email to licensel@sb-sib.ch).
-----
DR EMBL, Z81070; CAB03006.3; -.
DR EMBL, AF160248; AAD55518.1; -.
DR WormPep, F26E4.10; GE09694.
DR InterPro, IPR001159; DS_RBD.
DR Interpro, IPR000999; RNase_3.
DR Pfam, PF00035; dsrm; 1.
DR pfam, PF00636; Ribonuclease_3; 2.
DR SMART, SM00358; DSRM; 1.
DR SMART, SM00355; RIBoC; 2.
DR PROSITE, PS0137; RB.RD; 1.
DR PROSITE, PS00517; RNASE_3_1; 2.
DR PROSITE, PS0142; RNASE_3_2; 2.
KW Ribosome biogenesis; Hydrolase; Nuclease; Endonuclease; Repeat; RNA-binding; Nuclear protein.
FT DOMAIN 607 .. 781 RNASE III 1.
FT DOMAIN 833 .. 957 RNASE III 2.
FT FT FT DBM.
SQ DOMAIN 984 .. 1059 DRBM.
SEQUENCE 1086 AA; 125333 MW; 4AA7812DFBFB8 CRC64;

Query Match      19.3%, Score 1448; DB 1; Length 1086;
Best Local Similarity    31.2%; Pred.No. 5.7e-65;
Matches   364; Conservative     206; Mismatches 431; Indels 166; Gaps 31.
```

Dd	171	GDGSEFASADVSDSDSNS-----QDASITSEPTNNQAEADKT	208
Oy	414	DEDESSESCEDESDSTCS-----SSDSSEYFDVIAIETKRRKABDRH	519
Dd	209	GE-----VDEKOTCNRRNOQKARKRLNFEKEKOTLLAKKIDGRKTPNGIH	258
Oy	520	DEWYNDRPGQNNDRPLCKSCAKKARQCIIRISITRPGEAITKPCQPMNMGRLFN--RIT	577
Dd	259	PDISTNEGGLGNEBPERCERPEPIKGTGLKGYAGDKAIDCK--KSNENIHYTLART	316
Oy	578	VSPPTNFETDRPVYIECDHEEYIEFGESMFANAPL-----TNIPLCYIFINDIYIHFIE	633
Dd	317	PLPSENOUL--YRTIMALNGEFEEREGSLTTHNAPRLDCKTRAPICK--XSYMDYEOVAE	371
Oy	634	EMMP--ENFCVKGLEFSLFLFRDILLEYDMLKGPLFEEDSPCCRPFRHMPFRFBLPDG	692
Dd	372	EFMDECEDEPDCCMLEFEYIFHEIFEMLEDELKPKHIPSVDSECPMIHIMPEFYQ--TKMD	430
Oy	693	GKEYLSMHOILLYLL--RCSKALYDEEIAMLMOWELEMOKVAEEDKQAIYVNPCTKSS	751
Dd	431	LVQWMSKTYLATYFTSGSSEIIMSPREBVNMLCDAQIDQFTRNTSKIKOSIVLNTFKPSA	490
Oy	752	VRIOQLDRQENPDPVYIFPPIIIVFGIIRPAOLSYAGSPOFOKLMSYVYKRLHLANSPYK	811
Dd	491	IRAWFERDEDEKKEV-----YVNNAIIRAQYUUTIASLPRAFLEKTLINKKIO--EKSSGVY	545
Oy	812	QTDQOKLAOREEALOKITROKNTMRREYVLEYSOGFWKTGINSYDCQAHMMLPYLTHNR	871
Dd	546	NKDEK---TKNELEHLKRENNRSARMLKLEPAPAGELETGLKPDVAABHYVMTILACHNR	602
Oy	872	YHOCMLMDKILGTPDROCOLLOLAMTHPSHNLNPGMNDHARNSISNGCIQORPYG--D	929
Dd	603	YNFSLDVFEERYIEKFNDRVYETELALMHSSEKSHYGTPIDHKXMTYNGCYRR--KYGAED	661
Oy	930	RKVHNMHRKKGINTLLINISRLGODDPRTSRINNHRELEFLGDAVEFLTVSHLYULFR	989
Dd	662	KR-----EKKRVAGIMSLEFNIMKGTSGGER-----ILNHERLEYLODAVELYVSHHLFMYLT	714
Oy	990	SLEBGLATYKTAIVONOHMLAKKLELDPMYLAHGRDLCRESLDRIHAMANCEALIG	1049
Dd	715	HNFEGGLATYKTAIVYQNRNLATLAKNCRIDEMLOQSHGADLLINVAEFKHALANAEVAWA	774
Oy	1050	AVYLIEGSLSEAKOLFGLL--NDPDLREYVLANPLRLPLOLDERPNDROLIEFSPVLOKIT	1108
Dd	775	AIYLDGGIAPCDVIFESFAMYGHOIPVLKEKMDHINEHELKREDDPOGRDLSFTPLPSFTH	834
Oy	1109	EFEELAGIVFTHVWYLLNARFTLRTYVGNHLLTGLHNORMFELGSIIMOLVATEYLTTHPD	1166
Dd	835	ALBERLGIQGFNNITLLAKATTRNIRPNDLTKHNOBLEMDGSDVQLOLVSPFLRRFRPY	894
Oy	1169	HHECHTLRLRSSLVNNFTQAKVAEBELGOMERYATINDKTRPV---GLARTLTADLLESFT	1225
Dd	895	HHECHSMLRLTSLSVNOQTAVCCDDGEGTEFYI-----KAPYKTEPLKDKADADLEAFI	949
Oy	1226	AAITYDMDLEVYTHFMVNVCFPRPLKEITLNOQNDPKSLOLOCCCLTLR--TEKREDPIYX	1284
Dd	950	GALYVDKGIEICRAFIIRIVCPRLKHEITSEKKNADKSHLOWCMLAMRPSSSEBDMREY	1009
Oy	1285	KTLQTVPSHARTYTVAVYVYKGERIGGGKPSITQAEKMGAAAMDALDEKYNFPMOAHOKRFI	1344
Dd	1010	RYLDEIGPTNNRIRFKIAYVYKGRKLASABSNVHKAEIYALALAA--NLESMSFK---	1066
Oy	1345	GRKXYROELKEMRWERHOERPEDETD	1371
Dd	1065	-----MKAKNNSMFONMRRLEDDTSD	1086
RESULT 3			
RNC_BACSU	RNC_BACSU	STANDARD;	PRT; 249 AA.
AC	P51833; O31734;		
DT	01-OCN-1996 (Rel. 34, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease III (EC 3.1.26.3) (RNase III).
GN RNC OR RNCS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96257247; PubMed=8654983;
RA Oguro A., Kakeshita H., Takamatsu H., Nakamura K., Yamane K.;
RT "The effect of Srb, a homologue of the mammalian SRP receptor alpha-
RT subunit, on Bacillus subtilis growth and protein translocation.";
RT Gene 172:17-24(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNASE III DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D64116; BAA10976.1; -
CC EMBL: Z99112; CAB13466.1; -
CC Subtilast; BG11537; rnc.
CC InterPro: IPR001159; DS_RBD.
CC InterPro: IPR000999; RNase_3.
CC Pfam: PF00035; dsrm; 1.
CC Pfam: PF00636; Ribonuclease_3; 1.
CC SMART; SM00358; DSRM; 1.
CC DR SMART; SM00358; RIBOC; 1.
CC DR PROSITE; PS00137; DS_RBD; 1.
CC DR PROSITE; PS00517; RNASE_3_1; 1.
CC DR PROSITE; PS0142; RNASE_3_2; 1.
CC KM Hydroxylase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
CC FT DOMAIN 20 149 RNase III.
CC FT DOMAIN 226 242 DRBM.
CC FT CONFLICT 79 79 A -> P (IN REF. 1).
CC FT CONFLICT 171 171 S -> R (IN REF. 1).
CC FT CONFLICT 243 243 Q -> E (IN REF. 1).
CC FT CONFLICT 249 249 Q -> Q (IN REF. 1).
CC FT CONFLICT 249 249 Q -> Q (IN REF. 1).
CC SQ SEQUENCE 249 AA; 28425 MW; 304F3B0B5B7CCDC CRC64;

Query Match 4.2%; Score 315; DB 1; Length 249;
Best Local Similarity 34.7%; Pred. No. 2.3e-09;
Matches 85; Conservative 46; Mismatches 96; Indels 18; Gaps 7;

OY 1104 LQKLTFFERAGIVFTVHLLARAFLLRTVGFNHLTLGH--NORMEFLGDSIMOLVATEY 1161
DB 17 VQGFKEFORISVHQNQNELLAQAFTHSSVYVNEHKKRYEDERERLEFGDAVLELTISRF 76
OY 1162 LFIHPDHHEGHLTLRSLVNNRTQAKVAEELGMOEFAITNDKTKRPVG-LRIKTLADL 1220
DB 77 LFAKYPAWSEGDUTKLRAIVCEPSIVSLAHLSFGDVLVLKGHEMTGGRKRPALADV 136
OY 1221 LRSFIALYTDKDLKVYHNNVWCFEPRIKEILNODMNDPKSOLQCCILRTREGK--- 1277
DB 137 FEFALGATLALDGLFVESFLKVVPKINDGAFSHVM--DFRSQLOE---YVORDGKSL 192

```

```

OY 1278 EPDIPYKTLQTVGSPHARTYVAYFKGBRIGCGKPSIQOENKAMDALEKTNFPM 1337
DB 193 E-----YKISNEKGPANHREFEIVSLKGEPLGVGNRSKKEAEQHAQOALAKL---OK 244
OY 1338 AHOKR 1342
DB 245 HHTRQ 249

RESULT 4
RNC_VIRCH STANDARD; PRT; 225 AA.
ID NC_09KPB2;
AC 09KPB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease III (EC 3.1.26.3) (RNase III).
GN RNC OR VC2461.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR NI6961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "RNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -1- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNASE III DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE004316; AAF95603.1; -
CC TIGR: VC2461; -
CC InterPro: IPR001159; DS_RBD.
CC InterPro: IPR000999; RNase_3.
CC Pfam: PF00035; dsrm; 1.
CC Pfam: PF00636; Ribonuclease_3; 1.
CC SMART; SM00358; DSRM; 1.
CC DR SMART; SM00358; RIBOC; 1.
CC DR PROSITE; PS00137; DS_RBD; 1.
CC DR PROSITE; PS00517; RNASE_3_1; 1.
CC DR PROSITE; PS0142; RNASE_3_2; 1.
CC KM Hydroxylase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
CC FT DOMAIN 5 127 RNase III.
CC FT DOMAIN 154 224 DRBM.
CC SQ SEQUENCE 225 AA; 25010 MW; 692FEF4C580990C CRC64;

Query Match 3.8%; Score 285.5; DB 1; Length 225;
Best Local Similarity 32.2%; Pred. No. 6e-08;
Matches 75; Conservative 42; Mismatches 101; Indels 15; Gaps 5;

OY 1101 SPVIOKLTFFERAGIVFTVHLLARAFLLRTVGFNHLTLGHNRMEFLGDSIMOLVATE 1160
DB 1101 SPVIOKLTFFERAGIVFTVHLLARAFLLRTVGFNHLTLGHNRMEFLGDSIMOLVATE 1160

```



[illegible]

## RESULT 5

ID	RNC_PASMJ	STANDARD;	PRT;	225 AA.
AC	p57805;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Ribonuclease III (EC 3.1.26.3) (RNase III).			
GN	RNC OR PM0061.			
OS	Pasteurella multocida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae			
OC	Pasteurella.			
OX	NCBI_TaxID=747;			

RP SEQUENCE FROM N.A.  
 CC STRAIN=PM70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Pastian M.L., Whitam T.S., Kapur V.;  
 RT "Complete genome sequence of *Pasteurella multocida* Pm0.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING  
 CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: Endonuclease activity. Cleavage to 5'-  
 CC phosphonoester.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 DBRM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 RNASE III DOMAIN.

CC This SWI-PROT entry is copyright © it is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

Seq	Domain	154	224	DRBM	RNase III	Complete proteome
DR	EMBL: AEO06042; AAK02145.1; -					
DR	InterPro: IPR001159; DS_RBD; -					
DR	InterPro: IPR000999; RNase_3.					
DR	Pfam: PF00035; dsrm; 1					
DR	Pfam: PF00636; Ribonuclease_3; 1.					
DR	SMART: SM00358; DSRM; 1.					
DR	SMART: SM00353; RHOc; 1.					
DR	PROSITE: PS50137; DS_RBD; 1.					
DR	PROSITE: PS00517; RNase_3_1; 1.					
DR	PROSITE: PS50142; RNase_3_2; 1.					
KM	Hyalase; Nuclease; Endonuclease; RNA-binding; Complete proteome					
FT	DOMAIN	5	127			
FT	DOMAIN	154	224			
Seq	SEQUENCE	225 AA;	25461 MW;	Q02BACCA562B31C	CRC64;	

Query Match	3.8%;	Score 282;	DB 1;	Length 225;
Best Local Similarity	31.1%;	Pred. No. 8.9e-08;		
Matches 71;	Conservative 44;	Mismatches 101;	Indels 12;	Gaps 4

```

OY 1105 OKLIEEFAIGVIFTHVKLLARFTLRIVGNNHLIGHNORMEELGDSIMQVATEYLLFI 11
      | | : | | | : | | : | | : | | : | | : | | : | | : | | :
Db 3 QNTERLRQRIQGYQFNQDPA LKQALTHRSA-----AVKHNERLEELGDALINFTIAEALYH 57

```

[illegible]

## RESULT 6

ID	RNC_BACHD	STANDARD;	PRT;	263 AA.
AC	Q9KA05.			
DT	01-MAR-2002	(Rel. 4.1, Created)		
DT	01-MAR-2002	(Rel. 4.1, Last sequence update)		
DT	01-MAR-2002	(Rel. 4.1, Last annotation update)		
DE	Ribonuclease III (RC 3.1.26.3) (RNase III).			
GN	RNC OR RNCS OR BH2489.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group			
CC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=86665;			

RP SEQUENCE FROM N.A.  
RC STRAIN-C-125 / JCM 9153;  
RX MEDLINE-20512582; PubMed-11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fujii F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*  
RT halodurans and genomic sequence comparison with *Bacillus subtilis*.";   
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -1- FUNCTION: Digests double-stranded RNA. Involved in the processing  
CC of ribosomal RNA precursors and of some mRNAs (By similarity).  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
CC phosphomonoester.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC - SIMILARITY: CONTAINS 1 DRUG (DOUBLE STRANDED RNA BINDING) DOMAIN.  
CC - SIMILARITY: CONTAINS 1 RMASE III DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

or send an email to [license@slsb.sib.ch](mailto:license@slsb.sib.ch).

CC

DR EMBL, AP001515, BAB06208.1, -

DR InterPro, IPR001159, DS\_RBD

DR InterPro, IPR000999, RNase\_3

DR Pfam, PF00035, dsrm, 1

DR Pfam, PF00636, Ribonuclease\_3, 1

DR SMART, SM00358, DSRM, 1

DR SMART, SM00535, RIBOC, 1

DR PROSITE, PS00517, DS\_RBD, 1

DR PROSITE, PS00517, RNase\_3\_1, 1

DR PROSITE, PS01442, RNase\_3\_2, 1

DR Hydroxylase, Nuclease, RNA-binding, Complete proteome

FW DOMAIN 35 164 RNase III.

Query Match	3.8%; Score 282; DB 1; length 263;
SEQ SEQUENCE	263 AA; 30224 MW; 61B8E1C557CC5485 CRC64;

QY 1084 HP10LQEPNTDROLIETSPVLOKLEFEEBAICVIFTHVRLARAFTLRTVGNNH--LTIG 1141

```

Db      15 HSERRRPP---KRLFLTAQOQMFDELLRLTNLTFANKKLLVQAFTHSSYNEHRIQSC 71
      1142 HQQRNEFLGDSIMQVATVEYLFHPDHHGHLTLRSSLVNKKRQAVAEELSMQEFAT 1201
      Db      72 DNERLEFLGDAVLEAVSYLTKAEFQESGDMTKLRASIYCEPSLAQIAELHFGELV 131
      Qy      1202 TNDKRRPRYG-LRTKTLADLESFIALYTDKDLVEYHTEFMVNCPEPRLKEPILNQDMND 1260
      Db      132 LGKGEEMTKRRKRALADVFESFYGALXLDQMDAVILFERTYTPKISGASH-NMD 190
      Qy      1261 PPSQLQCCCLTLRTGKEKEDIPLYKTLQTVGSHARTYVAVYFGERIGCGKGPISQQA 1320
      Db      191 FKSQLOEF-----IORDNLGHIHYEIVQERGPANHREFVEVJLNNETLIGVGTGRSKKEA 245
      Qy      1321 EMGAMMDL 1329
      Db      246 EQHAAQAL 254

```

## RESULT 7

```

EXTN_TOBAC      STANDARD:      PRT:      620 AA.
ID      EXTN_TOBAC
AC      P13983;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
GN      HRCGN13.
OS      Nicotiana tabacum (Common tobacco).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC      Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.
OX      NCBI_TaxID=4097;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. XANTHT; TISSUE=leaf;
RA      MEDLINE=90128263; PubMed=2612909;
RX      Keller B., Lamb C.J.;
RT      "Specific expression of a novel cell wall hydroxyproline-rich
RT      glycoprotein gene in lateral root initiation.";
RL      Genes Dev. 3:1639-1646(1989).
CC      -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC      THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC      MAIN ROOT.
CC      -1- SUBCELLULAR LOCATION: Extracellular matrix.
CC      -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC      SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC      GLYCOSYLATED.

```

```

CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL, X13885; CAA32090.1; -.
CC      PIR, S06733; S06733.
CC      Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
CC      Hydroxylation.
CC      KW      SIGNAL
CC      FT      CHAIN      1      20      POTENTIAL.
CC      FT      REPEAT      70      73      EXTENSIN.
CC      FT      REPEAT      148      151      H-A-P-P.
CC      FT      DOMAIN      229      242      2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
CC      FT      REPEAT      229      235      1.
CC      FT      REPEAT      236      242      2.
CC      FT      DOMAIN      205      620      CONTAINS THE SER-PRO(4) REPEATS.
CC      FT      DOMAIN      499      600      3 X APPROXIMATE TANDEM REPEATS.
CC      FT      SEQUENCE      620      AA:      65406      MW:      641DD2278AB28524      CRC64;

```

Query Match 3.6%; Score 271.5; DB 1; Length 620;  
 Best Local Similarity 30.0%; Pred. No. 1e-06;  
 Matches 77; Conservative 19; Mismatches 80; Indels 81; Gaps 13;

```

Qy      29 PSADSPRPQMLRLHPQPPVQYQYEPSPASTTF-----NSPANEFLPPRPDVPFP 82
      Db      293 PPSPIYSP-----PPPASPSPPPPTPTFFSPPPAVSPPTYSPPPTYLPLFS 342
      Qy      83 -----PPMPSAGP-----LPQCP-----IRPFPMHOM-----RBP 110
      Db      343 SPITSPPPPVYSPPPPSYSPPPPTYLPPPPSSPPPSFSPPPTYEQSPPPPVAYSP 402
      Qy      111 FVPYPCPP-----MPPMPCPNNDPVG-APPQGT-----PFMMPP----- 149
      Db      403 LPAPPTYSPPPTYSPPPTTYAQPPPLPPTYSPPPVASPPPTYSPPPTYSPPPVAY 462
      Qy      150 SMPHPPPPVPMQOVNTQYRPGYSHNFPSPNSQNPNSSFLPSANSSSPHFLRP 209
      Db      463 AQPPPPPTYSPPPPAVSPPPSPISYSPPP-----QVQPLPPTFSPP-----PPRRILHP 514
      Qy      210 ----YPLKAPSERRSP 222
      Db      515 PPHQRPPTPTTGQPP 531

```

## RESULT 8

```

RNC_BUCAL      STANDARD:      PRT:      226 AA.
ID      RNC_BUCAL
AC      P57346;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Ribonuclease III (EC 3.1.26.3) (RNase III).
GN      RNC OR BU258.
OS      Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS      symbiotic bacterium).
OC      Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX      NCBI_TaxID=118099;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=TOKYO 1998;
RA      MEDLINE=20445173; PubMed=10993077;
RX      Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT      "Genome sequence of the endocellular bacterial symbiont of aphids
RT      Buchnera sp. APS.";
RL      Nature 407:81-86(2000).
CC      -1- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC      OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC      phosphomononucleotides.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC      -1- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC      -1- SIMILARITY: CONTAINS 1 RNASE III DOMAIN.

```

```

CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL, AP001118; BAB12968.1; -.
CC      InterPro: IPR001159; DS_RBD.
CC      InterPro: IPR000999; RNase_3.
CC      Pfam: PF00035; dsrm. 1.
CC      Pfam: PF00636; Ribonuclease_3; 1.
CC      SMART: SM00358; DSRW; 1.
CC      SMART: SM00535; RIBOC; 1.
CC      PROSITE: PS00137; DS_RBD; 1.
CC      PROSITE: PS00517; RNASE_3_1; 1.
CC      PROSITE: PS0142; RNASE_3_2; 1.
CC      Hydrolase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
CC      KW

```

FT DOMAIN 6 128 RNASE III.  
 FT DOMAIN 155 225 DRBM.  
 SQ SEQUENCE 226 AA; 25737 MW; F58DADF319A22133 CRC64;  
 Query Match 3.6%; Score 270; DB 1; Length 226;  
 Best Local Similarity 30.8%; Pred. No. 3.6e-07;  
 Matches 70; Conservative 47; Mismatches 94; Indels 16; Gaps 5;  
 QY 1109 EEEAAGVITFTVRLARAFTLRTVGFNHLTLGHNMREFLGDSTIMQVATYLFTHFPD 1168  
 DB 8 KIOXVIGYFTTHDOLKQALTHRSASK-----HNRLEFLDGLSIFVIAALVQHFFY 62  
 QY 1169 HHEGHITLRLSSLVNRTQAKVAEELGMOEYATINDKTRPVGLTKT-LADLLESFIA 1227  
 DB 63 IDEGDMSRKATLVNGTIAELAYEFDLGEYTLKGGEELKSGFRRESILANTVEALIGS 122  
 QY 1228 LVYDKLEVVHFMNVCFFPRLEKFLINDMDWPKSQLQCCLTLTBEKEPDIPYKTL 1287  
 DB 123 IYDSNIKTVEELIKWYERLEKISPGDTOKDPKRLDE---YLOSKHLISLPYFTV 177  
 QY 1288 QTVPSHARTYTVAVYKGERIG---CGKGPSIOQAEMGAADALEK 1331  
 DB 178 EYVGEAHNQLFT--HCKISTISEYILIGSSRRKAEODAAQKALIK 222

# RESULT 9 RNC\_LACIA ID RNC\_LACIA STANDARD; PRT; 231 AA.

AC 09CHD0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Ribonuclease III (EC 3.1.26.3) (RNase III).  
 GN RNC OR IL10802.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403.  
 RX MEDLINE=21335186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,  
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403";  
 RL Genome Res. 11:731-753(2001).  
 CC - FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING  
 CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS (BY SIMILARITY).  
 CC - CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
 CC phosphomonoester.  
 CC - SUBCELLULAR LOCATION: Cytoplasmic (By similarity)  
 CC - SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
 CC - SIMILARITY: CONTAINS 1 RNASE III DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE006313; AAK04900.1; -  
 DR InterPro: IPR001159; DS\_RBD.  
 DR InterPro: IPR000999; RNase\_3.  
 DR Pfam: PF00035; dsrm; 1.  
 DR Pfam: PF00636; Ribonuclease\_3; 1.  
 DR SMART: SM00358; DSRC; 1.  
 DR SMART: SM00535; RIBOC; 1.  
 DR PROSITE: PS50137; DS\_RBD; 1.  
 DR PROSITE: PS00517; RNASE\_3\_1; FALSE\_NEG.  
 DR PROSITE: PS50142; RNASE\_3\_2; 1.

KW Hydrolyase; Nuclease; Endonuclease; RNA-binding; Complete proteome.  
 FT DOMAIN 5 134 RNASE III.  
 FT DOMAIN 160 229 DRBM.  
 SQ SEQUENCE 231 AA; 26137 MW; 3D870B1F8B3AE916 CRC64;

Query Match 3.6%; Score 268.5; DB 1; Length 231;  
 Best Local Similarity 31.3%; Pred. No. 4.3e-07;  
 Matches 76; Conservative 48; Mismatches 88; Indels 31; Gaps 8;

QY 1104 LOKLEFEFAIGVFTVRLARAFTLRTVGFNHLTLGHNMREFLGDSTIMQVATYLFTHFPD 1154  
 DB 4 LOK-KLNNDYGLVFNDEDLTKTFT-----HSSFTNEERLPKIANNELEFLGVAL 54  
 QY 1155 QLVATEYFIHFPHDEHGLTLRLSSLVNRTQAKVAEELGMOEYATINDKTRPVGL-R 1213  
 DB 55 SLVDSIDYIKRYRPEKLEBELSKMSSIVRTSLANFSGCGFGLRLGHEEMGGDR 114  
 QY 1214 TKTADLLESFIALVYDKLEVVHFMNVCFFPRLEKFLINDMDWPKSQLQCCLT 1270  
 DB 115 ETTLENLEFEAFGLFIDQGMDEVAKFLQHVVIPHVK---NDQYVKYIDYKTELDE--- 167  
 QY 1271 TLRTGKRPDIPYKTLQTVGSHARTYTVAVYKGERIGCGKGPSIOQAEMGAADALE 1330  
 DB 168 -VLQIGGETTIS-YKILKEGPAHDRSFVAVFNNGELGKLGKSKVAEOKAENAIAIK 225  
 QY 1331 KYN 1333  
 DB 226 GQN 228

# RESULT 10 RNC\_ECOLI ID RNC\_ECOLI STANDARD; PRT; 226 AA.

AC P05797; P06141.  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ribonuclease III (EC 3.1.26.3) (RNase III).  
 GN RNC OR B2567 OR Z3848 OR ECS3433.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86039802; PubMed=3903434;  
 RA Nashimoto H., Uchida H.;  
 RT "DNA sequencing of the Escherichia coli ribonuclease III gene and its  
 RT mutations";  
 RL Mol. Gen. Genet. 201:25-29(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / SB221;  
 RX MEDLINE=85269601; PubMed=3895158;  
 RA March P.E., Ahn J., Inouye M.;  
 RT "The DNA sequence of the gene (rnc) encoding ribonuclease III of  
 RL Escherichia coli.";  
 RL Nucleic Acids Res. 13:4677-4685(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Nashimoto H., Saito N.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Sano Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";

RL Science.277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11205551;  
 RA Pena N.T., Plunkett G., Iii, Burtland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsudo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shimagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [7]  
 RP SEQUENCE OF 1-45 FROM N.A.  
 RX MEDLINE=89213943; PubMed=2540151;  
 RA Takiff H.E., Chen S.M., Court D.L.;  
 RT "Genetic analysis of the rnc operon of Escherichia coli.";  
 RL J. Bacteriol. 171:2581-2590(1989).  
 RN [8]  
 RP SEQUENCE OF 211-226 FROM N.A.  
 RX MEDLINE=87067411; PubMed=3097637;  
 RA Ahn J., March P.E., Takiff H.E., Inouye M.;  
 RT "A GTP-binding protein of Escherichia coli has homology to yeast Ras  
 RT proteins";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8849-8853(1986).  
 CC -1- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING  
 CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAs.  
 CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
 CC phosphomonoester.  
 CC -1- SUBUNIT: ORGANISED INTO A STRUCTURE (PROCESSOME) CONTAINING A  
 CC NUMBER OF RNA-PROCESSING ENZYMES.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: CONTAINS 1 DIRM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 RNASE III DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X02946; CAA26692.1; -;  
 DR EMBL: X02573; CAA26504.1; -;  
 DR EMBL: D64044; BAA10814.1; -;  
 DR EMBL: U36841; AAF79829.1; -;  
 DR EMBL: AE000343; AAC75620.1; -;  
 DR EMBL: AE005487; AAG57682.1; -;  
 DR EMBL: AP002562; BAB36856.1; -;  
 DR EMBL: M26415; AAA21843.1; -;  
 DR EMBL: M14558; AAA03241.1; -;  
 DR PIR: A26023; NREC3V.  
 DR PIR: A24022; NREC3V.  
 DR Ecogen: EG10857; rnc.  
 DR InterPro: IPR001159; DS\_RBD.  
 DR InterPro: IPR000999; RNase\_3.  
 DR Pfam: PF00035; dsrm.1.  
 DR Pfam: PF00636; Ribonuclease\_3; 1.  
 DR SMART: SM00358; DSRM; 1.  
 DR SMART: SM00535; RIBOC; 1.

```

DR PROSITE; PS50137; DS_RBD; 1.
DR PROSITE; PS00517; RNASE_3.1; 1.
DR PROSITE; PS50142; RNASE_3-2; 1.
KM HydroLase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
FT DOMAIN 6 128 RNASE III.
FT MUTAGEN 44 225 D-BM.
FT CONFLICT 168 195 G->D: LOSS OF ACTIVITY (MUTANT RNC-105).
FT FT HLPPLPTLVVQREAHDOEFTHCYVS -> PSAADLSG
SPGMSKRTIRNLSTARSV (IN REF. 2).
SO SEQUENCE 226 AA; 25550 MW; D9E2B58F2E0A3A5 CRC64;.

Query Match 3.5%; Score 262; DB 1; Length 226;
Best Local Similarity 29.9%; Pred. No. 8.9e-07;
Matches 70; Conservative %46; Mismatches 102; Indels 16; Gaps 6;

Oy 1101 SPVLQKLKEFEELGVITFHTRLLARAPFLTPVGNHMLTGHNORMEPFGISMOVATE 1160
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2 NFIV--IKRLQRKGTYTNHOELLQQALTHRSASK-----HNRLPEFGSIILSYIAN 54

Oy 1161 YLEFIHPDHGGHTLTLRSSLVNNRTOAKVAEEGMDEYATINDTKRPVGLRTKT-IAD 1219
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 55 ALYHFPRVDGDGDSMRBATLVRGNTILAELAREFELGECILRGCELGSGGFRRRESIIAD 114

Oy 1220 LLESTIALYTDKDYLEVYHTFMNVCFFRFLKAEFILNQDMNPKSQLOQCCLTLTEGERP 1279
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 115 TVEALIGGVFLSDIDQTVELKILLMWYQFRLEDISPGDKOKPKTRFLDE----YLQGRHL 169

Oy 1280 DIPFLKLOTGPSPSHARTYTVAAYFKG--ERIGCGKGPSIOQAEMGAAMDALK 1331
   :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 170 PLPLTVLVQVGGEAHDOEFTHCYVSGSLSEPV-VGTGSSRRKAQAAAEQALKK 222

RESULT 11
ZAP3_HUMAN STANDARD; PRT; 1822 AA.
AC ZAP3_HUMAN P49752; O9PIV7;
AC P49750; P49752; O9PIV7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3 (ZAP113).
DN ZAP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Moran A., Qin S., Abbasi N., Baradaran L., Birditt B.,
RA Bloom S., Doran M., Dickhoff R., Fleetwood P., Harrison G., James R.,
RA Kaur A., Madan A., Owen M.P., Ratcliffe A.J., Shaffer T., Hood L.;
RT "Sequencing of human chromosome 14q24.3 region."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Roggev E.L., Liang Y., Roggeva E.A., Levesque G.,
RA Ikeda M., Chi H., Iin C., Li G., Holman K., Tsuda T., Mar L.,
RA Fournet J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,
RA Plinesse L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Sansau P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L.,
RA Perlick-Vance M.A., Tanzi R.E., Rosses A.D., Fraser P.E.,
RA Rommens J.M., St George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
RT Alzheimer's disease."
RL Nature 375:754-760(1995).
CC -! SUBCELLULAR LOCATION: Nuclear (potential).
CC -! CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1661.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AC007956; AAF61275.1; -  
DR EMBL: LA0403; AAC42008.1; ALT\_FRAME.  
DR EMBL: LA0400; AAC42006.1; -  
KM Nuclear protein.

FT DOMAIN 15 205 PRO-RICH.  
FT DOMAIN 382 430 GLN-RICH.  
FT DOMAIN 807 1209 ARG-RICH.  
FT DOMAIN 1488 1577 ARG-RICH.  
FT CONFLICT 621 621 P -> S (IN REF. 2).  
FT CONFLICT 1404 1404 T -> I (IN REF. 2).  
FT CONFLICT 1821 1821 K -> E (IN REF. 2).  
SQ SEQUENCE 1822 AA; 204947 MW; 856CB83FE540C7D2 CRC64;

Query Match 3.5%; Score 260; DB 1; Length 1822;  
Best Local Similarity 20.6%; Pred. No. 1.4e-05;  
Matches 181; Conservative 103; Mismatches 327; Indels 268; Gaps 42;

QY 14 PGKACPRGRGHC---ARPSAP-----SFRQNL---RLHPOQPVQYQYEPSPAP 59  
DB 23 PVVLPASPGPYSSSTPAPSSSGFMSFREQHLAQLQLOMHQKQMCVQLDPHLLP 82  
QY 60 STFSNSPAPNPLPPR-----DFVPPRPPMSAGQYLPKPL-----RPPPNQ 106  
DB 83 -----PP- -LPPPVMGCGYGDQPPPPMP- - -PPGALSTYQKQKQYKHQ 127  
QY 107 MRHFPVPPCPMP 157  
DB 128 LHHQRDPPGLVPMELSP- -PESPVPVPGSYMPSSQSYMPSPPPPSY- -YPPSSQPYLP 185  
QY 158 PVPVQVNYVYRGYSHNPPPSFN-----SFQNNSSFLPSANNS-SSHPFH--- 206  
DB 186 PAQSP- -SQSPSPQS- -YLAPTPSYSSSSSSSSSYSSQSYLPSQASPSRPSQCHSKS 242  
QY 207 - -LPYPLPAKPSERSPERLKYDHRHNDHSHGGERHRSIDRRERGSPDRRRQDSK 264  
DB 243 QLLAP- -PPPSAPRPNKTYYQOELES-----CAKKKSTEQQAAPERPSTW--- 288  
QY 265 YRSYDYGRTPSRRRSYERSRE- -RERERHNRNRNRSPLERSYKREKSGRSGYL 320  
DB 289 -----TPROQOQYWRQHLLSLQQRKVALPGHKKGPPVAKOTPEVYKEVYVPAF 339  
QY 321 SYVPE- -PAGTPRLPGELIKNTDSMAPRLIVNHRSPREKKARWEEDKRWSDQSSG 379  
DB 340 SQVPESSPSEEPPLP- -PPNTEVPPPLP- - - - - - - - - - - - - - - - - 365  
QY 380 KDKNYSIKKEPEETMPDKNEEEBELKPVWIRCTHSENYSSDPMQV- - - - - - - - 436  
DB 366 -----PEE- -POSEDEPEEARLKOLQAAAH- - - - - - - - - - - - - - - 407  
QY 437 GTSRLDLVDKFEELGSGROEKAARPPWERPKTLDELESSSECESEDSTCSSS 496  
DB 408 KHTQLQOIIQOYOOIOPRPHIQATTPRPIIP- - - - - - - - - - - - - - - - 462  
QY 497 SDESEVD- 548  
DB 463 OSSQVPEKPPRALLPPVSGSAPPTT- - - - - - - - - - - - - - - - - - 515  
QY 549 HSIYP- 596  
DB 516 YSFSFSSDQGLGESSAPSGQITAVKDMPVASGGL- - - - - - - - - - - - - - 569  
QY 597 HEYIEGFSFMAHAPLTNPPLCKVIRFNIDYTHIIEEMPENVCVKGDELPSLFLRDI 656  
DB 570 PRR-FEDIGSRCEGP- 583  
QY 657 LELYDNMLKLPFEDSPPCPRFHMPRVYRLPDGKE- - - - - - - - - - - - - - 713  
DB 584 -----RPKGPREFGNRPDGR- 633

QY 714 VPBEELANLQMELE- 770  
DB 634 SPRO- 681  
QY 771 IIVHEGIRPAQLSYAGDPQYQKLMKSYVLRHLIANSRK 809  
DB 682 -----FKMQSAFSTIADVDVAAQSNENLSDSQDEPK 716

RESULT 12  
ID GPL\_CHLRE STANDARD; PRF; 555 AA.  
AC 09PRO6: 003927;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Vegetative cell wall protein gpl precursor (Hydroxyproline-rich glycoprotein 1).  
GN Gpl  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=11258910;  
RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kitz S., Drees J., Goodenough U.W.;  
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in plant hydroxyproline-rich glycoproteins.";  
RL Biochemistry 40:2978-2987(2001).  
RN [2]  
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.  
RX MEDLINE=91017504; PubMed=1699225;  
RA Adair W.S., Apt K.E.;  
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).  
CC - - - - - FUNCTION: Major component of the outer cell wall W6 (crystalline) layer.  
CC - - - - - SUBUNIT: Associates with GP2 and GP3.  
CC - - - - - PTM: N-glycosylated and O-glycosylated.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF309494; AAC45420.1; -  
DR EMBL: M58496; AAC69706.1; ALT\_SEQ.  
DR GlycoSuiteDB: Q9PPO6; -  
DR InterPro: IPR003882; Pistil\_extensin.  
DR InterPro: IPR002965; P-rich\_extensin.  
DR PRINTS: PR01217; PRICHEXTENSIN.  
DR PRINTS: PR01218; PSTLEXTENSIN.  
KM Glycoprotein; Repeat; signal.  
FT SIGNAL 1 29  
FT CHAIN 30 555  
FT DOMAIN 40 339  
FT DOMAIN 259 279  
FT CAROXYD 399 399  
FT CAROXYD 455 455  
FT CAROXYD 493 493  
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502P5 CRC64;

Query Match 3.5%; Score 259.5; DB 1; Length 555;  
Best Local Similarity 33.7%; Pred. No. 3.5e-06;  
Matches 82; Conservative 13; Mismatches 101; Indels 47; Gaps 10;







D	b	851	VSXKGS-POGRRLLOQOMRATKFNVLTTTYYEVIKDKAVLAIQKYMIIIDEGHBMKNNH	909
Q	y	774	-----HFGRPQOLSTAGDPQOYKIMSYKILRHLLA-----NSPKVK	811
D	b	910	CKLTQVNLNTHY-IAPRRLLTGTPLOKNCPELIMALLNELLNFIKSCSTFEQWENAPAT	968
Q	y	812	QTDOKQKLAQEEELQKIRQKNRREPVYVELSSQSGWKIGISDVCSQAHMIPVLTTHIR	871
D	b	969	TGEVLENEETLL-----IIR-----LHVLPFLRLRK	1000
Q	y	872	---YHOCIMLIDLKIGYEQ-DKCLTOLMTHPSHHLNGMNPDHARNSLSCGIRPKY	927
D	b	1001	KEYEHQ---LPKVERIIXKQMSALQ-----	1023
Q	y	928	GDRKVHHMHRKKGINTLINMSRLGODDPSPSRINHNREFLGDVAVELTSHLYL	987
D	b	1024	---RVLTKHMQSGV-LLTDSSEK-----HG-----	1047
Q	y	988	FPSLEBGLATYTAIYONHMLAKKLELDPMLYAH-----GPDLC	1031
D	b	1048	-----KGAALMNTIYO-----LRKLCNHPFM-FOHIEEKYCDHTGGHGVSGPDLY	1094
D	b	1095	RVS-----GRPE-----LDRIL-----PKKA-----	1112
Q	y	1092	NTDROLIETSPVLOKLTNEEBAIG-VITFHVRL-----LARAFTLTGFMHLT	1139
D	b	1113	---TNRVRLPFCQMOCMTIEDYLGWROFGYLRLLDGTPTKAEKRGELLRKRNK- GSDV	1169
Q	y	1140	LGNHOMREPLGDSIMQIVATEYLFHPDPHBSGHLTLRSSLVNNNTQAKAAEELGMOEY	1199
D	b	1170	FLSTRAGGGLN-LQTAADVLFEDSDMNPBODLQADRAHRIGORNEVR-----	1219
Q	y	1200	AITNDTKRRRVGRTKTLADILESFLAA-----LYTDKLEVYHTTMNCCFRLKEF-----	1252
D	b	1220	-----RLMTVNSVERIILAAARVKNIMDEKVIQAGMDQSGTSGEROOFLQ	1267
Q	y	1253	ILNDMDPKRSOLO---OCCLTRLTEGKEBPDIPLVYLTQVPSHARTVTVAVYKRGRI	1303
D	b	1268	ILHQDNEEBEENBVPDDENIMNIMARSBEELIEFRMDAERKK-----EDEEI	1316
Q	y	1310	GCGKPSIQQAEMGAAM---DALEKYNFPQAHQKRTIGRKYRQ-----ELKEMR	1357
D	b	1317	HPGERRLIDSESLPDMWLTKKDDVEVERFHY-QYDEDTILIGRSRORKEVDYDLSLTEREM	1374
Q	y	1358	-----EREHOREPDEDIEDIKK	1374
D	b	1375	LKAIDGAEFDEEBEDSDSKRRKKR	1399

RESULT	15		
RNC_SALTY			
ID	RNC_SALTY	STANDARD:	PRT: 226 AA.
AC	Q56056;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-MAR-2002 (Rel. 41, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Ribonuclease III (EC 3.1.26.3) (RNase III).		
CN	RNC OR STM2581.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Salmonella.		
OX	NCBI_TaxId=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=TSM117;		
RX	MDLINE=97295239; PubMed=9150881;		
RA	Anderson P.E., Matsunaga J., Simons E.L., Simons R.W.;		
RT	"structure and regulation of the Salmonella typhimurium rnc- <i>era</i> - <i>reco</i>		
RL	operon.";		
RL	Biochimie 78:1025-1034(1996).		
RN	[2]		

RP	SEQUENCE FROM N.A.
RC	STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX	MEDLINE=21534948; PubMed=11677609;
RA	McClelland M., Sanderson K.E., Speich J., Clifton S.W., Latreille P.,
RA	Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA	Waterston R., Wilson R.K.;
RT	Complete genome sequence of <i>Salmonella enterica</i> serovar Typhimurium
RL	LT2.
RL	Nature 413:852-856(2001).
CC	- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC	OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAS.
CC	- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC	phosphonooxester
CC	- SUBUNIT: ORGANIZED INTO A STRUCTURE (PROCESSOME) CONTAINING A
CC	NUMBER OF RNA-PROCESSING ENZYMES.
CC	- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC	- SIMILARITY: CONTAINS 1 DRAEM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC	- SIMILARITY: CONTAINS 1 RNASE III DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on lists
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).
CC	-----
DR	EMBL; U048415; AAA92440.1; -.
DR	EMBL; AE008817; BAL21475.1; -.
DR	StyGene; SC10695; rnc.
DR	InterPro; IPR001159; DS_RBD.
DR	InterPro; IPR000999; RNase_3.
DR	Pfam; PF00035; dsrm; 1.
DR	Pfam; PF00636; Ribonuclease_3; 1.
DR	SMART; SM00358; DSRM; 1.
DR	SMART; SM00535; RIBOC; 1.
DR	PROSITE; PS00137; DS_RBD; 1.
DR	PROSITE; PS00517; RNASE_3.1; 1.
DR	PROSITE; PS0142; RNASE_3.2; 1.
KW	Hydrolase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
FT	DOMAIN 6 128
FT	DOMAIN 208 224
FT	DOMAIN 57 58
FT	CONFLICT 73 74 AT -> S (IN REF. 1).
FT	CONFLICT 217 220 AT -> DSV (IN REF. 1).
FT	CONFLICT 217 220 BOAL -> NSV (IN REF. 1).
FT	SEQUENCE 226 AA; 25505 MW; 5752C0113C0A055A CRC64;

Search completed: October 9, 2002, 18:24:01  
Job time : 41 secs

Oy	1101	S P V O K I T E B E A L G Y I F H V A U L L A R A P T K R V G N H L T G S H O R M E J D S I M O U V A N E	1160
Dz	2	N P I V - - I N R L Q R K L G T F N H O E I L Q O A L T H R S A S S K - - - - H N E R L E F L D S I L S E V I A N	54
Oy	1161	Y L F I H P D H E G H L T L R S S I V N N R T O A K V A E E L G M O E A Y T I N D K T R P G L T K T - L A D	1219
Dz	55	A L Y H R P R V D E G S M S K R A T V Y R G T I A M E L A R E P L G S C L R L G E L K S G F P R E S I L A D	114
Oy	1220	L E S F I A L Y T D K D E Y V H T F M N V C P P R L K E F I L N O D M D P K S O L O O C C L T L R T E G K E P	1279
Dz	115	T V E A L I G V F L D I N I O T V E Q L I L N Y M K T R L D E I S P G D K O X P K T R L O E - - - - Y L O G R H L	169
Oy	1280	D I P L K T L Q V G S H A R T V T A V Y R K G - - E R I G C K G R S I O A A M G A M A L E K	1331
Dz	170	P L S Y L V A V O Y R G A H D O E F T I C O V S G S E P V - V O T G S S R K A O A A A E O A L K	222

```
Search completed: October 9, 2002, 18:24:01
Job time : 41 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 17:13:02 ; Search time 66 Seconds  
(without alignments)  
2312.354 Million cell updates/sec

Title: US-09-900-425a-2

Perfect score: 7500  
Sequence: 1 MMQNGTCHMSEHFHGRGCPDR.....MRWERHQREPFDETDIKK 1374

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4071	54.3	769	22	AAB92635	Human protein sequ
2	2963	39.5	1337	22	ABB58539	Drosophila melano
3	1961	26.1	378	22	AAU20587	Human secreted pro
4	1961	26.1	378	22	AAU21744	Human secreted pro
5	1956	26.1	378	22	AAU20385	Novel human neopla
6	1320	17.6	263	22	AAU21580	Human secreted pro
7	1295.5	17.3	301	22	AAU21580	Novel human neopla
8	1289	17.2	267	22	AAU21580	Human breast cance
9	1122	15.0	277	22	AAU21580	Human breast cance
10	496	6.6	115	20	AAU21580	Human breast cance
11	496	6.6	115	21	AAU21580	Human 5' EST secre
						Human secreted pro

12	340	4.5	1151	22	ABB61598	Drosophila melano
13	311	4.1	560	22	ABG21040	Novel human diagno
14	303.5	4.0	406	22	ABG27250	Novel human diagno
15	277.5	3.7	232	20	AAW81354	S. pneumoniae rnc
16	277	3.7	1537	22	ABB66989	Drosophila melano
17	276.5	3.7	1278	22	AAU39259	Human polypeptide
18	275.5	3.7	594	22	ABB61362	Drosophila melano
19	273	3.6	973	22	AAW41045	Human polypeptide
20	272.5	3.6	702	22	ABB63064	Drosophila melano
21	267	3.6	243	20	AAU09366	Staphylococcus aur
22	265.5	3.5	470	22	ABG21932	Novel human diagno
23	261	3.5	572	18	AAW31855	Mycobacterium tube
24	261	3.5	763	18	AAW31852	S. epidermidis ope
25	260.5	3.5	245	22	AAU83282	Drosophila melano
26	259.5	3.5	1638	22	ABB61946	Novel human diagno
27	259	3.5	1217	22	ABG09876	Novel human diagno
28	258	3.4	598	22	ABG14000	Drosophila melano
29	255.5	3.4	707	22	ABB68526	Drosophila melano
30	254	3.4	2249	22	ABB60977	Drosophila melano
31	254	3.4	2249	22	AAU09769	Drosophila melano
32	253	3.4	1239	22	ABG09877	Novel human diagno
33	245	3.3	565	22	AAW78694	Human protein seq
34	244.5	3.3	446	22	ABB70063	Drosophila melano
35	242	3.2	255	18	AAU20772	H. pylori cytoplas
36	240	3.2	615	21	AAU46152	Arabidopsis thalia
37	240	3.2	679	21	AAU46151	Arabidopsis thalia
38	240	3.2	691	21	AAU46150	Sugar beet chitina
39	239.5	3.2	439	13	AAU28150	Neospora NC-p65 ve
40	239.5	3.2	861	22	ABB03068	Human expressed po
41	238	3.2	691	22	AAU23180	Novel human diagno
42	236.5	3.2	325	22	ABG21919	Novel human diagno
43	236	3.1	1030	22	ABG23699	Novel human diagno
44	236	3.1	299	12	AAU13995	Rape extensin. Br
45	235.5	3.1	299	12	AAU13995	

#### ALIGNMENTS

RESULT 1	
ID	AAU2635 standard; Protein; 769 AA.
XX	
AC	AAU2635;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human protein sequence SEQ ID NO:10949.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
XX	
PR	27-AUG-1999; 99JP-0300253.
XX	
PR	11-JAN-2000; 2000JP-0118776.
XX	
PR	02-MAY-2000; 2000JP-0183767.
XX	
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto U;
XX	
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI; 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection  
PR and/or diagnosis of the abnormality of the proteins encoded by the  
PS full-length cDNAs -  
XX Claim 8; SEQ ID 10949; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH1629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX Sequence 769 AA:  
SQ  
Query Match 54.3%; Score 4071; DB 22; Length 769;  
Best Local Similarity 99.5%; Pred. No. 1.9e-280;  
Matches 765; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY . 606 MFAHAPLNIPLCKVIRINIDYTHFIEEMPEFCVAGLEFSFLERDLLEYLDMWK 665  
DB 1 MFAHAPLNIPLCKVIRINIDYTHFIEEMPEFCVAGLEFSFLERDLLEYLDMWK 60  
QY 666 GLFEDSPCCRFHMFHMFVFLPDGKEVLSMHQILLYLRCSKALVPEEETANMLQW 725  
DB 61 GLFEDSPCCRFHMFHMFVFLPDGKEVLSMHQILLYLRCSKALVPEEETANMLQW 120  
QY 726 EELEWQKYAECKGMIVTNPSTKPSVRIDQREQNPDIYTPPIIVHSGIRPAQSYA 785  
DB 121 EELEWQKYAECKGMIVTNPSTKPSVRIDQREQNPDIYTPPIIVHSGIRPAQSYA 180  
QY 786 GPPQYQKLMKSYVKRLHLANSFKYQTDKQKLAQREBLQKIRQKNTMRREVVELSSQ 845  
DB 181 GPPQYQKLMKSYVKRLHLANSFKYQTDKQKLAQREBLQKIRQKNTMRREVVELSSQ 240  
QY 846 GPMKTGIRSDVCOHMMPLVLTNHIYRQCLMHLDKLIGYTFODCLQLQMLTPSHSLN 905  
DB 241 GPMKTGIRSDVCOHMMPLVLTNHIYRQCLMHLDKLIGYTFODCLQLQMLTPSHSLN 300  
QY 906 FGMNDPHANSLNSGCIROPKYGDKKVHHMMKKGINTLINMSRLQDDPTPSRINHN 965  
DB 301 FGMNDPHANSLNSGCIROPKYGDKKVHHMMKKGINTLINMSRLQDDPTPSRINHN 360  
QY 966 ERLERPLGDAVVEFLTSVNLHYLFPSLEEGGLATYRTATVQONHMLAKKLELDFEMLYA 1025  
DB 361 ERLERPLGDAVVEFLTSVNLHYLFPSLEEGGLATYRTATVQONHMLAKKLELDFEMLYA 420  
QY 1026 HGPDLCSRESDLRHANANCFEALIGAVYLEGSLEAKQIFGRLLFNDPDLREVNLVPLHP 1085  
DB 421 HGPDLCSRESDLRHANANCFEALIGAVYLEGSLEAKQIFGRLLFNDPDLREVNLVPLHP 480  
QY 1086 LQLOEPNTDROLIETSPVLQKLTFFEEAIGVITFHVRLARAFLLRTVGFNHLTGNQR 1145  
DB 481 LQLOEPNTDROLIETSPVLQKLTFFEEAIGVITFHVRLARAFLLRTVGFNHLTGNQR 540  
QY 1146 MEFLGDSIMQVATERYLTFHPDHHGHGHTLLRSSLVNNRQAKVAEELGMEVATITDK 1205  
DB 1205 MEFLGDSIMQVATERYLTFHPDHHGHGHTLLRSSLVNNRQAKVAEELGMEVATITDK 1205

DB 541 MEFLGDSIMQVATERYLTFHPDHHGHGHTLLRSSLVNNRQAKVAEELGMEVATITDK 600  
QY 1206 TKRPVGLRTKTLADLESFIALYTDKDEYVHTFPMNVCFFPRLKEFILNDPMNPKSQL 1265  
DB 601 TKRPVGLRTKTLADLESFIALYTDKDEYVHTFPMNVCFFPRLKEFILNDPMNPKSQL 660  
QY 1266 QQCCLTLRTGKRPDIPLYKTLQTVGSHARTYVAVYFKGRIKCGKPSIQAEKGA 1325  
DB 661 QQCCLTLRTGKRPDIPLYKTLQTVGSHARTYVAVYFKGRIKCGKPSIQAEKGA 720  
QY 1326 MDALERYNPQMAHQKRFGRKYROELKEMRWEREHOEPEDETDIK 1374  
DB 721 MDALERYNPQMAHQKRFGRKYROELKEMRWEREHOEPEDETDIK 769  
RESULT 2  
ID ABB58539 standard; Protein; 1327 AA.  
XX  
AC ABB58539;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 2409.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PMD, Myers EM;  
XX  
DR WPI; 2001-656860/75.  
XX  
DR N-PSDB; ABL02642.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 2409; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
CC  
XX  
XX Sequence 1327 AA:  
SQ  
Query Match 39.5%; Score 2963; DB 22; Length 1327;  
Best Local Similarity 45.6%; Pred. No. 2e-201;  
Matches 621; Conservative 213; Mismatches 327; Indels 200; Gaps 32;  
QY 90 GGPLPCCIPPPPNHOMRHPFVPPCFPPMPMPCPN-NPVPGPAP-----P 137  
DB 3 QPPLPPPPVQPA-----PPPPPPPEEDLSPGCVGVPHNYSNNSHSHQ 46

2

Qy	138	GQGFPPMPPSPHPRRPPVMDQVNOUYOPGY-----SHHNPP	179
Db	47	SSKLDYVYETAPYASS---VPSTDPQPPAYIEGYATVNEQAQYGGQESHYQYQY	103
Qy	180	PSFNSSONNPSSELPSPANSSSPHFRHLPYPL---PAPSEKRSRPERLKHUYDHRD	235
Db	104	PA-----SSSFLYES-----YKPPDRPAYSSNYPSPBQ-----RTS	139
Qy	236	HSRGRGRHSLDRENGRGRPDRRRODSKRSYDNGRTFSPNRSYENSREPERRHNH	295
Db	140	NSSSQGHHPV-----GYSSGR-----RYEORHDQ-----BHRQDSRYAHNEPHGHY	183
Qy	296	DNRRSPLEKSYKKEKRSGRSYGLSVPEPACCTELPGEILKTKMDWAPLEIVNRS	355
Db	184	AHRQAKSQSGY---YQSAARN-----QVSDQSPGHHERERN	219
Qy	356	PSRKKKA---RWEBKDR---WSDNQSGKDKNYSIKEKEPEETMDKNEEBELIK	409
Db	220	ETLEKTAKEKRYETERRLLRWQSNFC-----EKPEYVKKMNLSEADAPV	267
Qy	410	PWYRCHSENYSSDPMQDQGVTVGTSRLDLDKFEELGSGROEKAKARPPWP	469
Db	268	ESWRRSPALYIERK---SENEYRGARLQCLTLPDEBELLORAKVRKRLPYVPP	323
Qy	470	KTKL-----DEDESSSESECEDESDSTCSSSSSEVFDV---IAIKRKKAH	514
Db	324	PKRARREVCXKHKHSEACSSSSSDDDSDEDA-----FKTEQDCMELSRKVOH	373
Qy	515	PDRJHDLANDNQOMNDGCLCKCSAKARTGRHSITYGEALIRCPMRTNNARLHY	574
Db	374	PORHAAHMHNDGEMNDGPLCRCSAKSRRTGRHGYGELGYKLCODNSNACKRLHY	433
Qy	575	RIVSPPTNLTLPRTVTEVDEHETFEQGSMPAHPLNIPJCKVYIRPNIDYTHFTEE	634
Db	434	RISISPTNLTLPRTIKKDEHEFLPEGSULSHVRLSDLPVCVYIRNIEYITIEYEE	493
Qy	635	MPRNEFCVKGLEFLFPLNDLIELYDWNLCPL---FEDSPCCPREHMPRVPLP	690
Db	494	KMPNPTIHELDIFKFLFHELEVDVFNMPNLPNGVNEES---CPAHFEPREVRDLP	550
Qy	691	DGGEVFSMHQILLYLRCALVPEEBETANMLQMELMQKUYAEKQCMYTNGTGRPS	750
Db	551	DNGEVFLAMEVRLYLLDANSQALVEBQOILLHNOJSQSMQNYDFIKGMLYTKKGYPC	610
Qy	751	SVRIDLDREQNDV-----TTPVIYHFGIRPALSAYADPOYOKLMSYVKLR	801
Db	611	SLRVDQDRLN---NSDRECVDERETGSHALVHFGICHPOLSYAPNEQYKAMREYKX	668
Qy	802	HLNLSKPVQOTOKOXAQNEEALOKIROKNMRRRVUTELSQQGFWKGTISDVCOHAM	861
Db	669	HLNMNSKPSFKDRKLEEKQRIQKQKROGRRKRNITVAISSEGYRFGJMGCDVQOAM	728
Qy	862	MLPVLNHIRYHOCIMLIDLTYETQODRCLQDANTHSHNLNFGMNDHARNSLSCNG	921
Db	729	LIPVLGHLRFHKSJDLLEESIGYRKRNRIQLQALTHPSYKEMGNTGNDHARNSLTCG	788
Qy	922	IRQKYQDCKRVHNNHMKKQINFLNMSLQODPTSRIRNNRLEFLGDVAYEGLTS	981
Db	789	IRQPEYQDCKRIHYNNKRIINTLVISMSFGKEHETVSNITNNRLEFLGDVAYEFLSS	848
Qy	982	VHLVYLRPSLEEGILAVYRAIVONOHAMAKRLDLDFMYLYANGRDLGRSDDRHMA	1041
Db	849	IHLFEMPELEBGLATYRAIYONOHALLAKKLDLDEBFMLYAHBSDCHELELRHMA	908
Qy	1042	NCFEALIGAVYLEGSLAEAKOLFGRLLF---NDPDLREWMLNPLNPLQLOEPNTDQLEET	1100
Db	909	NCFEALIGALLDGGCIYVADEVFTDALFRQDEKLISIMKLNREHPLQDEBPLGDSCTDS	968
Qy	1101	SPVLQKTLFEERELIYVETFYVLLAFAFLTKRYGHNHLTGNQMRFLGDSIMQVATE	1166
Db	969	YVLIKEJLKEEDSDIGIKFKIRILLARAFDRSIRGFHNLTSNQRIEFGDVTQDLOLISE	1028

QY	1161	YLTFHPDHHGHGHTLTLSRLSVNNPFOAKYAEELGMOEYAI-TNKTTRPGLRPTKLTAD	1219
Db	1029	YLRHHPFHHHGHSLTLRSLSVNNKRTQAAVCDLGMPTKPAIVANK-----ADLTKTKRAD	1084
QY	1220	LLESEFIATLYDKDLEEVHFMVAFCEFPRLKEPILNQMDNDRKSLOOCCLTLTPT-EGKE	1278
Db	1085	LLEAFGLATLYDKGLXLCDFCHVCLFPPRLQLEINMQMDNDRKSLQOCCLTLTRMOSGE	1144
QY	1279	PDIPLYYTLQTVGSPSHARTYVAIVYKRGERIGCGKPSIQQAEMGAMDALEKYN--FPQ	1336
Db	1145	PDIPYKYVKEASGPTNTRVYKVAIVYFRSKRLATSSGSSSIQQAEMNAKQAALENSRLPEQ	1204
QY	1337	MAHOKRFLGRKYRQEL---KEMWEEHQRERDEPDEDIKK	1374
Db	1205	LDHOKRVIAKSIKQGTGENLNDSDROHE-----EKIKR	1239
RESULT 3			
ID	AAU20587	standard; Protein; 378 AA.	
XX	AAU20587;		
XX			
DT	04-DEC-2001	(first entry)	
XX			
DE		Human secreted protein, Seq ID No 579.	
XX			
XX		Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;	
KW		rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;	
KW		cytoprotective; thrombolytic; antimicrobial; ophthalmological;	
KW		cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;	
KW		multiple sclerosis; cancer; hyperproliferative disorder; infection;	
KW		Gaucher's disease; neurological disease; cerebrovascular disorder;	
KW		thrombosis; wound healing.	
XX			
OS		Homo sapiens.	
XX			
PM		W0200155326-A2.	
XX			
PD		02-AUG-2001.	
XX			
PE		17-JAN-2001; 2001WO-US01347.	
XX			
PR		31-JAN-2000; 2000US-0179065.	
XX			
PA		(HUMA-) HUMAN GENOME SCI INC.	
XX			
PI		Rosen CA, Barash SC, Ruben SM;	
XX			
DR		WPI; 2001-451931/48.	
XX		N-PSDB; AAS33296.	
PT		New nucleic acids and polypeptides, useful for diagnosing, preventing	
PT		or treating medical conditions -	
PS		Claim 11; SEQ ID No 579; 753bp; English.	
XX			
CC	The invention relates to novel isolated nucleic acid molecules (I)		
CC	encoding human secreted proteins (II). (I) and (II) are used to prevent,		
CC	treat or ameliorate a medical condition in e.g. humans, mice, rabbits, in		
CC	goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used, in		
CC	the prevention, treatment and diagnosis of diseases associated with		
CC	inappropriate expression of secreted proteins. (I) and complementary		
CC	sequences may also be used as DNA probes in diagnostic assays (e.g.		
CC	polymerase chain reactions (PCR) to detect and quantitate the presence		
CC	of similar nucleic acid sequences in samples, and so which patients may		
CC	be in need of restorative therapy. (II) may also be used as antigens in		
CC	the production of antibodies and in assays to identify modulators		
CC	(agonists and antagonists) of the expression and activity of the secreted		
CC	proteins. The anti-(II) antibodies and antagonists may also be used to		
CC	down regulate expression and activity of (II). The anti-(III) antibodies		
CC	may also be used as diagnostic agents for detecting the presence of (II)		
CC	in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The		

disorders include for example: immune/autoimmune diseases (e.g. HIV  
(human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
melanoma, neoplasms of the breast or liver, Sezary syndrome and  
Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/  
cerebrovascular disorders (e.g. cardiac arrest, tachycardia, and  
angina and thrombosis), infections caused by bacteria, viruses and  
fungi and ocular disorders (e.g. corneal infections). (I) and (II),  
agonists, antagonists and antibodies can also be used to promote wound  
healing, maintain organs before transplantation, and support cell culture  
of primary tissues. AAU20342-AAU20666 represent human secreted protein  
amino acid sequences, and related sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification but was obtained in electronic format directly from WIPO  
at: ftp.wipo.int/pub/publ/published\_pct\_sequences.

Sequence 378 AA;

Query Match 26.1%; Score 1961; DB 22; Length 378;  
Best Local Similarity 98.9%; Pred. No. 5.3e-131;  
Matches 374; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 997 ATYRAIVQNHLMALAKKLELDPPMLYAHGPDLCRESDLRHANANCFEALIGAVYEGS 1056  
Dy 1 ATYRAIVQNHLMALAKKLELDPPMLYAHGPDLCRESDLRHANANCFEALIGAVYEGS 60

Qy 1057 LEEAKQLEGRLLFNDPDLREVMYLPPLHLOEPNTDROLIETSPYLOKITEPEAAIGV 1116  
Dy 61 LEEAKQLEGRLLFNDPDLREVMYLPPLHLOEPNTDROLIETSPYLOKITEPEAAIGV 120

Qy 1117 IFTHVRLARAFYLTATVGFNHLITLGHNRMEFLDSIMQVATETYLTHPDDHGHGHTL 1176  
Dy 121 IFTHVRLARAFYLTATVGFNHLITLGHNRMEFLDSIMQVATETYLTHPDDHGHGHTL 180

Qy 1177 LRSSLVNRRQAKVAEEIGMOEYAITNDKTRPGYLRKTADLLESPFIALYDKDLEY 1236  
Dy 181 LRSSLVNRRQAKVAEEIGMOEYAITNDKTRPGYLRKTADLLESPFIALYDKDLEY 240

Qy 1237 VHTFMNVCFFPRLKEFILNQDMNDPKSLOQCCLTLRTEGKEPDIPYKTLQTVGPPSHAR 1296  
Dy 241 VHTFMNVCFFPRLKEFILNQDMNDPKSLOQCCLTLRTEGKEPDIPYKTLQTVGPPSHAR 300

Qy 1297 TYTVAVYFKGERIGCGKSPSTIOQAEMGAAMDALKEKYNPOMAHOKRFTGRYRDELKEMR 1356  
Dy 301 TYTVAVYFKGERIGCGKSPSTIOQAEMGAAMDALKEKYNPOMAHOKRFTGRYRDELKEMR 360

Qy 1357 WEREHOREPDETDIKK 1374  
Dy 361 WEREHOREPDETDIKK 378

RESULT 4  
AAU21744  
ID AAU21744 standard; Protein: 378 AA.  
XX  
AC AAU21744;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Novel human neoplastic disease associated polypeptide #177.  
XX  
KW Human; neoplastic disease associated polypeptide; cancer;  
KW hyperproliferative disorder; neural disorder; immune system disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
KW neuroprotective; cytostatic; anti inflammatory; vasotropic.  
XX  
OS Homo sapiens.  
XX  
PN WO200155163-A1.  
XX  
PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01358.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465558/50.  
DR N-PSDB; AAS34943.

XX Novel polypeptides and polynucleotides useful as diagnostic reagents to  
PI diagnose diseases or disorders associated with aberrant expression or  
PI activity of polypeptides, and for treating cancers, Rheumatoid  
PI arthritis  
PS Claim 11; SEQ ID NO 471; 687pp; English.  
XX  
CC The present invention relates to the isolation of novel human neoplastic  
CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA  
CC sequences encoding for these polypeptides. The sequences of the  
CC invention are useful in the diagnosis, treatment, prevention and/or  
CC prognosis of disorders involving neoplastic disease such as  
CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder  
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar  
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may  
CC also be useful for treating other disorders such as neural disorders,  
CC immune system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
CC and renal disorders. The polynucleotide sequences of the invention are  
CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human  
CC neoplastic disease associated polypeptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 378 AA:  
Query Match 26.1%; Score 1961; DB 22; Length 378;  
Best Local Similarity 98.9%; Pred. No. 5.3e-11;  
Matches 374; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 997 ATYRTAIYQNOHMLAKKLEIDPMLYAHGPDLCSRDLEHMANCFEALIGAVYLEGS 1056  
DB 1 ATYRTAIYQNOHMLAKKLEIDPMLYAHGPDLCSRDLEHMANCFEALIGAVYLEGS 60  
QY 1057 LEEAKOLGRLLEFNDPREWLNYPPLHPLQLOEPNTROLIETSPYOKLTPEEBAIGV 1116  
DB 61 LEEAKOLGRLLEFNDPREWLNYPPLHPLQLOEPNTROLIETSPYOKLTPEEBAIGV 120  
QY 1117 IFTHVRLARAFETLTGVSFNHLTGHNOEMFLDLSIMQVATERYLTFHPDHHGHJTL 1176  
DB 121 IFTHVRLARAFETLTGVSFNHLTGHNOEMFLDLSIMQVATERYLTFHPDHHGHJTL 180  
QY 1177 LRSSLVNNRTOAKAAVEELGMOEYATITNDKTRPVGLKTKTLADLESFIAALYTDKOLEY 1236  
DB 181 LRSSLVNNRTOAKAAVEELGMOEYATITNDKTRPVGLKTKTLADLESFIAALYTDKOLEY 240  
QY 1237 VHTFNNVCFPPPLKEFTLINOBNPDKSLOQCCLTLRREGKEPDIPLYKTLQTVGPSHAR 1296  
DB 241 VHTFNNVCFPPPLKEFTLINOBNPDKSLOQCCLTLRREGKEPDIPLYKTLQTVGPSHAR 300  
QY 1297 TYTVAVYFKGERIGCGKPSIQQAEMGAADALEKYNPQAHOKRFTIGKRYROELKEMR 1356  
DB 301 TYTVAVYFKGERIGCGKPSIQQAEMGAADALEKYNPQAHOKRFTIGKRYROELKEMR 360  
QY 1357 WEREHOEREPPDETEDIK 1374  
DB 361 WEREHOEREPPDETEDIK 378  
RESULT 5  
AAU20385  
ID AAU20385 standard; Protein; 378 AA.  
XX  
XX AAU20385;  
XX  
XX 06-DEC-2001 (first entry)  
XX  
XX Human secreted protein. Seq ID No 377.  
DE  
XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;  
KW rheumatoid arthritis; antarteriosclerotic; cardiac; vascular;

KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;  
KW cytosolic; Alzheimer's disease; Parkinson's disease; human; cancer;  
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;  
KW Gaucher's disease; neurological disease; cerebrovascular disorder;  
KW thrombosis; wound healing.  
XX  
OS Homo sapiens.  
XX  
PN WO200155326-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01347.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI: 2001-451931/48.  
DR N-PSDB; AAS33094.  
XX  
PT New nucleic acids and polypeptides, useful for diagnosing, preventing  
PT or treating medical conditions -  
XX  
PS Claim 11; SEQ ID No 377; 753pp; English.  
XX  
CC The invention relates to novel isolated nucleic acid molecules (I)  
CC encoding human secreted proteins (II). (I) and (II) are used to prevent,  
CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,  
CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in  
CC the prevention, treatment and diagnosis of diseases associated with  
CC inappropriate expression of secreted proteins. (I) and complementary  
CC sequences may also be used as DNA probes in diagnostic assays (e.g.  
CC polymerase chain reactions (PCR)) to detect and quantitate the presence  
CC of similar nucleic acid sequences (II) in samples, and so which patients may  
CC be in need of restorative therapy. (II) may also be used as antigens in  
CC the production of antibodies and in assays to identify modulators  
CC (agonists and antagonists) of the expression and activity of the secreted  
CC proteins. The anti-(II) antibodies and antagonists may also be used to  
CC down regulate expression and activity of (II). The anti-(II) antibodies  
CC may also be used as diagnostic agents for detecting the presence of (II)  
CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The  
CC disorders include for example: immune/autoimmune diseases (e.g. HIV  
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/  
CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,  
CC angina and thrombosis), infections caused by bacteria, viruses and  
CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),  
CC agonists, antagonists and antibodies can also be used to promote wound  
CC healing, maintain organs before transplantation, and support cell culture  
CC of primary tissues. AA020342-AA020666 represent human secreted protein  
CC amino acid sequences, and related sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO  
CC at: ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 378 AA;  
XX  
Query Match 26.1%; Score 1956; DB 22; Length 378;  
Best Local Similarity 98.7%; Pred. No. 1,2e-110;  
Matches 373; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 997 ATYRTAIVQNHILAMIAKLEIDPEPMAYAHGPDLCRESDLRHANANCFEALIGAVYLEGS 1056  
DB 1 ATYRTAIVQNHILAMIAKLEIDPEPMAYAHGPDLCRESDLRHANANCFEALIGAVYLEGS 60  
QY 1057 LEEAKQLFGRLFLNDPDLREVWLNTPLEHPLQLQEPNTDRLIETSPVLOKLEFEBAIGV 1116  
LEAKQLFGRLFLNDPDLREVWLNTPLEHPLQLQEPNTDRLIETSPVLOKLEFEBAIGV 1116

DB 61 LEEAKQLFGRLFLNDPDLREVWLNTPLEHPLQLQEPNTDRLIETSPVLOKLEFEBAIGV 120  
QY 1117 IFTHVLLARAFLLRTVGFNHLTLGHNRMEPLGDSIMQVATEYLTFHPDHGCHLTL 1176  
DB 121 IFTHVLLARAFLLRTVGFNHLTLGHNRMEPLGDSIMQVATEYLTFHPDHGCHLTL 180  
QY 1177 LRSILVNNRTQAKVAEELQMEYAITNDKTRPVGLRRTKTLADLLESTFALAYTPDKLEY 1236  
DB 181 LRSILVNNRTQAKVAEELQMEYAITNDKTRPVGLRRTKTLADLLESTFALAYTPDKLEY 240  
QY 1237 VHTFMNVCFEPRLKEFTLQMDNNDPSQLQCCCLTLRTGKEPDIPLYKTLQTVGSPSHAR 1296  
DB 241 VHTFMNVCFEPRLKEFTLQMDNNDPSQLQCCCLTLRTGKEPDIPLYKTLQTVGSPSHAR 300  
QY 1297 TYTVAVYFKEGIRGCGKPSIQAEKGAMDALEKYNFPMAHORFIRGKRYQELKEMR 1356  
DB 301 TYTVAVYFKEGIRGCGKPSIQAEKGAMDALEKYNFPMAHORFIRGKRYQELKEMR 360  
QY 1357 WEREHOERPEDETIKK 1374  
DB 361 WEREHOERPEDETIKK 378  
RESULT 6  
AA021580  
ID AAU21580 standard; Protein; 263 AA.  
XX  
AC AAU21580;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Novel human neoplastic disease associated polypeptide #13.  
XX  
KW Human: neoplastic disease associated polypeptide; cancer;  
KW hyperproliferative disorder; neural disorder; immune system disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
KW neuroprotective; cytosolic; anti inflammatory; vasotropic.  
XX  
OS Homo sapiens.  
XX  
PN WO200155163-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01358.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214386.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.





	Best Local Similarity	97.7%;	Pred. No. 1,1e-85;	
	Matches 251;	Conservative	0;	Mismatches 6; Indels 0; Gaps 0;
QY	1118	FTFVRLARAFRLTRTGFNHLTLFGNQMEFLGDSIMOLVATEYLFIHFDPHHEGHLL	1177	
Db	7	FYSOCLLARAFRLTRTGFNHLTLFGNQMEFLGDSIMOLVATEYLFIHFDPHHEGHLL	66	
QY	1178	RSSLVNNRTOAKYAEELGMOEYAITNDKTRPVGRLRTKLADLESTIAALVYDKDLEV	1237	
Db	67	RSSLVNNRTOAKYAEELGMOEYAITNDKTRPVALRTKLADLESTIAALVYDKDLEV	126	
QY	1238	HTFMNVCFFPRLEKFLINDMDNDPKSOLOOCCFLTRTGKGRPDLPYLKTLQTVGSPHART	1297	
Db	127	HTFMNVCFFPRLEKFLINDMDNDPKSOLOOCCFLTRTGKGRPDLPYLKTLQTVGSPHART	186	
QY	1298	YTYAVYFCKGRIGCGKSPISIOQAEKGAMDALDEKYNFPMOHAKRFYGRKYROELKEMEN	1357	
Db	187	YTYAVYFCKGRIGCGKSPISIOQAEKGAMDALDEKYNFPMOHAKRFYGRKYROELKEMEN	246	
QY	1358	EREHOREPDETEDIKK	1374	
Db	247	EREHOREPDETEDIKK	263	

XX	RESISTIT	7
XX	AAB63281	
XX	AAB63281 standard; Protein; 301 AA.	
XX		
XX	AAB63281;	
XX		
XX	26-MAR-2001 (first entry)	
XX		
XX	Human breast cancer associated antigen protein sequence SEQ ID NO:643.	
XX		
XX	Human; breast cancer; gastric cancer; prostate cancer; diagnosis;	
XX	cancer associated antigen; cytostatic; cancer vaccine.	
XX		
XX	Homo sapiens.	
XX		
XX	WO200073801-A2.	
XX		
XX	07-DEC-2000.	
XX		
XX	26-MAY-2000; 2000WO-US14749.	
XX		
XX	28-MAY-1999; 99US-0136526.	
XX	10-SEP-1999; 99US-0153454.	
XX		
XX	(LUDW-) LUDWIG INST CANCER RES.	
XX		
XX	Obata Y;	
XX		
XX	WPL; 2001-025274/03.	
XX		
XX	Nucleic acids encoding breast, gastric and prostate cancer associated	
XX	antigen precursors, useful for diagnosing and treating a condition	
XX	characterized by expression of an abnormal amount of a protein, e.g.	
XX	cancer -	
XX		
XX	Example 1; Page 499; 799pp; English.	
XX		
XX	AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014	
XX	represent nucleotide sequences encoding human breast, gastric and	
XX	prostate cancer associated antigen precursors (CAAP) respectively.	
XX	AAF63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970	
XX	represent human breast, gastric and prostate CAAP protein sequence	
XX	respectively. CAAPs have cytostatic activity and can be used in the	
XX	production of cancer vaccines. The human CAAP proteins, peptides, nucleic	
XX	acids or anti-CAAP antibodies are useful for diagnosing and treating a	
XX	condition characterised by expression of an abnormal amount of a protein,	
XX	e.g. cancer.	
XX		
XX	Sequence 301 AA;	
XX		

	Query Match	17.3%;	Score 1295.5;	DB 22;	Length 301;
	Best Local Similarity	89.7%;	Pred. No. 7.2e-84;		
	Matches 252;	Conservative	4;	Mismatches 18;	Indels 7;
					Gaps 4.
QY	AMMEPVLTHTHTIRHOCMLMDKLTGCTFODRCLLOLATMTHPSHHLNGMNDHARNSLSN	919			
Db	15 AMMLPVLTHTHTIRHOCMLMDKLTGCTFODRCLLOLATMTHPSHHLNGMNDHARNSLSN	74			
QY	920 CGIRQPKYGRKXHHNMRRKKGINTLTINMSRIGQDDTPSRINHNERLEFGDAVVEFL	979			
Db	75 CGIRQPKYGRKXHHNMRRKKGINTLTINMSRIGQDDTPSRINHNERLEFGDAVVEFL	134			
QY	980 TSVHLYLTPSLEEGGIATYRTAIYONQHLAMAKLELDFEMLYANGPDLCRESDLRHA	1039			
Db	135 TSVHLYLTPSLEEGGIATYRTAIYONQHLAMAKLELDFEMLYANGPDLCRESDLRHA	194			
QY	1040 MANCFEALIGAVYIEGSLSEBAKOLFGRLLPNDPDLREYVWLYPLHPIQLQDEPNTDRLT-	1098			
Db	195 MANCFEALIGAVYIEGSLSEBAKOLFGRLLPNDPDLREYVWLYPLHPIQLQDEPNTDRLIG	254			
QY	1099 ETSVPLOKLTFFE-EAIGVITFTVRLTLAAFLTRVGVNHL	1138			
Db	255 NEFQFYKKLTFFERNLDGVNFL--LIGSDFX-AXGSHL	290			

RESULT 8	
AA663379	
ID	AA663379 standard; Protein; 267 AA.
XX	
AC	AA663379;
XX	
DT	26-MAR-2001 (first entry)
XX	
DE	Human breast cancer associated antigen protein sequence SEQ ID NO:741.
XX	
KW	Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX	
KW	cancer associated antigen; cytostatic; cancer vaccine.
XX	
OS	Homo sapiens.
XX	
PN	MO200073801-A2.
XX	
PD	07-DEC-2000.
XX	
PF	26-MAY-2000; 2000MO-US14749.
XX	
PR	28-MAY-1999; 99US-0136526.
XX	
PR	10-SEP-1999; 99US-0153454.
XX	
PA	(LUDW-) LUDWIG INST CANCER RES.
XX	
PI	Obata Y;
XX	
DR	WPI; 2001-025274/03.
XX	
PT	Nucleic acids encoding breast, gastric and prostate cancer associated
XX	
PT	antigen precursors, useful for diagnosing and treating a condition
XX	
PT	characterized by expression of an abnormal amount of a protein, e.g.
XX	
XX	cancer -
XX	
PS	Example 1; Page 543; 799pp; English.
XX	
PS	AA6622422 to AA6622626, AA6622627 to AA6622773 and AA6622774 to AA6623014
XX	
CC	represent nucleotide sequences encoding human breast, gastric and
XX	
CC	prostate cancer associated antigen precursors (CAAP) respectively.
XX	
CC	AA663332 to AA663467, AA663468 to AA663721 and AA663722 to AA663970
XX	
CC	represent human breast, gastric and prostate CAAP protein sequence
XX	
CC	respectively. CAAPs have cytostatic activity and can be used in the
XX	
CC	production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX	
CC	acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX	
CC	condition characterised by expression of an abnormal amount of a protein,
XX	
CC	e.g. cancer.

XX Sequence 267 AA; 17.2%; Score 1289; DB 22; Length 267;  
Query Match Best Local Similarity 94.6%; Pred. No. 1.8e-83;  
Matches 246; Conservative 2; Mismatches 10; Indels 2; Gaps 2;

QY 861 MMPLVLTNHRHQCLMHLDKLGYTFODRCILQLAMTHPSNHLNCGMPDHARNSLNC 920  
DB 1 MMPLVLTNHRHQCLMHLDKLGYTFODRCILQLAMTHPSNHLNCGMPDHARNSLNC 60  
QY 921 GIRQPKYGRKRVHNMRRKGINTLINIMSRIGQDDPTPSRINHNRELFGLDAVVEFLT 980  
DB 61 GIRQPKYGRKRVHNMRRKGINTLINIMSRIGQDDPTPSRINHNRELFGLDAVVEFLT 120  
QY 981 SVHLTYLFPSEGLATYRTAVQNOHLAMAKKLELDRFMYAHGPDLCSRESDLRHAM 1040  
DB 121 SVHLTYLFPSEGLATYRTAVQNOHLAMAKKLELDRFMYAHGPDLCSRESDLRHAM 180  
QY 1041 ANCFEALIGAVYLEGSLEAKOLFGRLLFNDPDLREVWLNYPRLPQLQDEPNTDROLI-E 1099  
DB 181 ANCFEALIGAVYLEGSLEAKOLFGRLLFNDPDLREVWLNYPRLPQLQDEPNTDROLI-G 240  
QY 1100 TSPVLQKLEFE-EAIGVIF 1118  
DB 241 FFOFYKKLTFEERNQLGVNE 260

RESULT 9  
AAB63383  
ID AAB63383 standard; Protein; 277 AA.  
AC AAB63383;  
DT 26-MAR-2001 (first entry)  
DE Human breast cancer associated antigen protein sequence SEQ ID NO:745.  
XX  
XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;  
KW cancer associated antigen; cytostatic; cancer vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200073801-A2.  
XX  
PD 07-DEC-2000.  
XX  
XX 26-MAY-2000; 2000WO-US14749.  
XX  
XX 28-MAY-1999; 99US-0136526.  
PR 10-SEP-1999; 99US-0153454.  
XX  
XX (LUDW-) LUDWIG INST CANCER RES.  
XX  
XX Obata Y;  
XX  
XX WPI; 2001-025274/03.  
XX  
XX Nucleic acids encoding breast, gastric and prostate cancer associated  
PT antigen precursors, useful for diagnosing and treating a condition  
PT characterized by expression of an abnormal amount of a protein, e.g.  
PT cancer -  
XX  
XX Example 1; Page 545; 799pp; English.  
XX  
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014  
CC represent nucleotide sequences encoding human breast, gastric and  
CC prostate cancer associated antigen precursors (CAAP), respectively.  
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970  
CC represent human breast, gastric and prostate CAAP protein sequence  
CC respectively. CAAPs have cytostatic activity and can be used in the  
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic  
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a

CC condition characterised by expression of an abnormal amount of a protein,  
CC e.g. cancer.  
XX  
XX Sequence 277 AA; 15.0%; Score 1122; DB 22; Length 277;  
Query Match Best Local Similarity 83.1%; Pred. No. 1.4e-71;  
Matches 231; Conservative 6; Mismatches 21; Indels 20; Gaps 4;

QY 942 INTLINIMSRIGQDDPTPSRINHNRELFGLDAVVEFLTSVHLTYLFPSEGLATYRT 1001  
DB 1 INTLINIMSRIGQDDPTPSRINHNRELFGLDAVVEFLTSVHLTYLFPSEGLATYRT 60  
QY 1002 AIVQNOHLAMAKKLELDRFMYAHGPDLCSRESDLRHAMANCFEALIGAVYLEGSLEAK 1061  
DB 61 AIVQNOHLAMAKKLELDRFMYAHGPDLCSRESDLRHAMANCFEALIGAVYLEGSLEAK 120  
QY 1062 QLFGRLLFNDPDLREVWLNYPRLPQLQDEPNTDROLIETSPVLQKLEFEFAIGVIFTHV 1121  
DB 121 QLFGRLLFNDPDLREVWLNYPRLPQLQDEPNTDROLIETSPVLQKLEFEFAIGVIFTHV 180  
QY 1122 RLIAAFILRTYGFVHNLTLGNQRMFEIGDSIMOLV-ATEYLFIHFPHHGHITLLRSS 1180  
DB 181 RLIAAFILRTYGFVHNLTLGNQRMFEIGDSIMOLVPOSTVFTS-QIITGHILFY--- 236  
QY 1181 LVNNRTQAKVABELGMEYAITNDKTRPVGLRTFTLA 1218  
DB 237 -----AKASLGE-----INKNFRPKGNKRAMA 259

RESULT 10  
AAV12224  
ID AAV12224 standard; Protein; 115 AA.  
XX  
XX AAV12224;  
XX  
XX 18-JUN-1999 (first entry)  
XX  
XX  
XX Human 5' EST secreted protein SEQ ID NO: 537.  
XX  
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemokine; chemokine; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
XX  
XX  
OS Homo sapiens.  
XX  
XX WO9906554-A2.  
PN  
XX 11-FEB-1999.  
XX  
XX 31-JUL-1998; 98WO-IB01238.  
XX  
XX 01-AUG-1997; 97US-0905134.  
XX  
XX (GEST ) GENSET.  
XX  
XX Ducleit A, Dumas Milne Edwards J, Lacroix B;  
XX  
XX WPI; 1999-153784/13.  
XX  
XX N-PSDB: AAX41057.  
XX  
XX New nucleic acids encoding human secreted proteins - obtained from  
PT CDNA libraries prepared from kidney, fetal kidney, dystrophic  
PT muscle, muscle and heart tissue  
XX  
XX Claim 34; Page 599; 622pp; English.  
XX  
XX AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins, and encode the proteins given in AAV101602 and



SQ Sequence 1151 AA;

Query Match	4.5%;	Score 340;	DB 22;	Length 1151;
Best Local Similarity	22.5%;	Pred. No. 3.1e-15;		
Matches 173;	Conservative 83;	Mismatches 280;	Indels 234;	Gaps 35

QY	28	RSASASRPOULRLHQQOQRPVQYQVEPPASPTTENSANPFL-----PPRPDYVPP	82
Dp	362	RPASBPKEVP-----PPAPGVESPFGOPAPSPRFPDPPTHTIEPPPAEPTLVPP	415
QY	83	PPMPSAOGPLPCP--IRPPFPHQARHPFVPCPPMPMPMPCN--NPVPGPAP--	136
Dp	416	PPAPPTIKPPPPAPPTVEEPP-----PPAPPTVEPPPPPAFTKIEPPPPAPAE	469
QY	137	--PGQTFPPMPPSPMHP--PPVMPQOQNNQYPP-----GY-----	172
Dp	470	VEPPPPAPATELEPPPPAPPEKVLPPPPAPKAEALTPRRAKSGNGFITELATPKES	529
QY	173	-----SHHNF-----PPSPFSFQNP-----SSFLPSANSSSPH	203
Dp	530	ETROKPINVNTYKEDQOPATTEDEKODKPLGLGADAKAVSTASESTESTSTSSOS--	587
QY	204	FRLHPYPLPKAPSERSPRLKHYDHRHRDHSHGGRHSTLDR--RESGRSPDRROD	262
Dp	588	-----KSHKSDKEKEKDRHRHRSDKH-----RRSTDRDRSRKSHKSHS	633
QY	263	SRYSRDYDRGR--TPSRHRSYERSERERERHRDRNRPSLER-----SYKK	309
Dp	-634	SSSSSKHSSSSSSSKHKSSSSKNDKSSSSSPHNSSSSSKRSQTTSSRHSSSHKK	693
QY	310	-----EKKRSGRSGLSVPEPPACPTTELGEIITKNTDSNAPLELYNHNSPREKKRAR	364
Dp	694	HKSSSSSSRSERDQG-----KEKREKRSQSRKS-----HHSSSSSSSSSR	736
QY	365	WEESKDRWSD--NOSSGSKDN-----YSIKE-----KEPETMPDKNE	401
Dp	737	KDHRGGDRDORHNSKNTSGSANKAIHDDHSEKEXYQRRGSSQSNDEGRPPSSGGRAKNS	796
QY	402	EEEE--LKKPWIRCTHSENNYSDDPMQVGS--TVGTSRLDLYDKFEELGSR	455
Dp	797	QPEDSAATMSDAPAVENANGTNGNSNGSTAGACDQNVGVIYVSDILOQSTSSFEVLTAGS	856
QY	456	QEKAKARPWEPPKPTLDDLESSSECESEDSTCSSSD-----SEVDVIAEIK	509
Dp	857	QSHDRASSKHHEPIDGKGRADQNPETELEAKODE--CASONVPPYEPOTLADSVDPLE	915
QY	510	RKKHAPRLHDELWYNDPQGMNDPLOCCKSAKARCTGIRISIVGEEAIPCAPPTNNMG	569
Dp	916	AKES-MOTLEDE-----KVTWEEN--KSEESKP-----	943
QY	570	RLHFRTVSPPTNLDRPTVIEYDDHEXIFEGFSMFAHAPLTNIPLC--KVIRPNIDY	627
Dp	944	-----ENPPECVDEPAQVGDVEDP-----PQDADKPAPTVPVINSQSDSESADP	988
QY	628	TIHIEEMPEBNCVKGLELFSLEFLDILLEYDW-----NKGPLEED	671
Dp	989	VTRH-ENTDE-----FETRIOLINOLIEDRKNLNRSSD	1023

### RESULT 1.3

ID ABG21040 standard; Protein; 560 AA.

AC ABG21040;

DT 18-FEB-2002 (first entry)

Novel human diagnostic protein #21031.

KW Human, chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX  
PN WO200175067-A2.

PD 11-OCT-2001

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC

PI Drmanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.

DR N-PSDB; AASB

PR New isolated polynucleotide and encoded polypeptides, useful in  
PR diagnostics, forensics, gene mapping, identification of mutations  
PR responsible for genetic disorders or other traits and to assess  
PR biodiversity -

PS Claim 20; SEQ ID NO 51399; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABC00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp://wipo.int/pub/published\\_pat\\_sequences](ftp://wipo.int/pub/published_pat_sequences).

Sequence 560 AA;

Query Match	4.18;	Score 311;	DB 22;	Length 560;
-------------	-------	------------	--------	-------------

Best Local Similarity 29.5%; Pred. No. 1.4e-13;

Matches 114; Conservative 22; Mismatches 131; Indels 120; Gaps 20;

Oy	15	GRGCRGR-----GGHGARSAPSFRRQULRLHQOAPRVQYQYPRSPAPSTTFSNSP	67
Db	203	GRG--RGRRPGAVPGDGRPGRPAGGCR-----PPRPAQGTTPRRPLGTSFGCRP	256
Oy	68	APNELPRAPDVPV-----PRMP-----PSAQ--GRUPCSDTPRRPNQMR	108
Db	251	GRPCRRPGQVLRPLRPLAGPRNRGDRPRRVLFPRQGRPGQPLGRLPGR--PRVPGVG-P	308
Oy	109	HRPVRPCFRPMPRRPMPCRNP-----PVRGA-PRCQGTFRPMP	147
Db	309	PRGPRPRQGGPRPRGPRPRRGTGSLGHPLLAPPSASPRGPRPGGAPRPSHVNAFRP	368
Oy	148	PRSMRNP-----PRPMPOOVVQYQYPCGSHNFRPP-----SFSNQN	187
Db	369	PRTNCGMPTSDSRGRRPRDPR--YGRPRPRYDRGDYGRPGKEMDTARTPLSEAFEEIHN	424
Oy	188	N-----PSSEFLSANSSSPH-----FRHLPRYLPRAPSERKSRPELKH	227
Db	425	RNRATSSAISRAVSDASGAVDGSALPTLTVALTSLKTTPKPY----ADDRCKTFLISS	480
Oy	228	YDD--HHRDHSHGEGENHRGLDRRERGRSPDRRRODSRYSDVDGRGTFSPRHNYSER	285



CC presence of the protein; or (b) by detecting nucleic acid encoding the  
CC protein. Ab are useful as antibacterial agents; to isolate or identify  
CC rnc-expressing clones and for affinity purification. Fragments of the DNA  
CC are useful as probes or primers to isolate full-length or related  
CC sequences; to screen for drugs, and to diagnose or stage infections, also  
CC for genotyping and serotyping of infective agents (e.g. by detecting  
CC mutations).

XX  
SQ Sequence 232 AA;

Query Match 3.7%; Score 277.5; DB 20; Length 232;  
Best Local Similarity 32.5%; Pred. No. 1e-11;  
Matches 75; Conservative 47; Mismatches 92; Indels 17; Gaps 8;

QY 1108 TEEFEAIGVIFTHVRLARAFILRTYGFNH--LTIGHNORMEFLGDSIMQLVATEYLFTH 1165  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 6 TVLKNHFAIEFADKNLETAFTHTSYANHRLLKISHNERLEFLGDAVLQLISEYLYKK 65  
QY 1166 FPDHHEGHLTLRSSLVNNRTOAKVAEELGMOEYAITNDKTRPVGLRTR--TLADLES 1223  
:  
Db 66 YPKRPEGDLSKLAMIVREESLAGFARDCQDFQF--IKLGKGEKSGGRNRDITLGDAPFA 124  
QY 1224 FIALYTDKDELYVHTFPMVCEFFRLK--EFLINDMNDPKSLOQCCLTLTGKKEPDI 1281  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Db 125 FIGALLLDKDAKKEFIYQVMIPKVEAGEFEMITDY--KTHLE--LLQVNG---DV 175  
QY 1282 PL-YKTLQTVGSPSHARTYTVAVYFKGERICGKSPISQOAEKGAAMDALAK 1331  
:  
Db -176 AIRYOVISSEGRANDKVFDEYVVEGKSTIGCGGRSKTIAEQEAAKNAYEK 226

Search completed: October 9, 2002, 18:23:12  
Job time : 72 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 18:22:18 ; Search time 34 Seconds  
(without alignments)  
987,081 Million cell updates/sec

Title: US-09-900-425A-2

Perfect score: 7500  
Sequence: 1 MMGGTCHRMSPHNGRCPR.....MRERERHOREDEDEDIDK 1374

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277.5	3.7	232	US-08-869-674-2	Sequence 2, Appl
2	277.5	3.7	232	US-09-213-010-2	Sequence 2, Appl
3	267	3.6	243	US-08-965-087-2	Sequence 2, Appl
4	226	3.0	334	5202236-3	Patent No. 5202236
5	226	3.0	1274	US-09-095-443-2	Sequence 2, Appl
6	225.5	3.0	331	5202236-37	Patent No. 5202236
7	220.5	2.9	1185	US-09-041-886-23	Sequence 23, Appl
8	218	2.9	306	US-08-217-327-6	Sequence 6, Appl
9	216	2.9	905	US-08-574-959A-9	Sequence 9, Appl
10	216	2.9	905	US-09-357-014-9	Sequence 9, Appl
11	216	2.9	1135	US-08-574-959A-7	Sequence 7, Appl
12	216	2.9	1135	US-09-357-014-7	Sequence 7, Appl
13	203	2.7	1255	US-09-080-897-4	Sequence 4, Appl
14	203	2.7	1255	US-08-899-595-1	Sequence 1, Appl
15	203	2.7	1255	US-09-323-735-4	Sequence 4, Appl
16	202	2.7	174	US-08-818-112-143	Sequence 143, App
17	202	2.7	174	US-08-818-111-138	Sequence 138, App
18	202	2.7	174	US-09-056-556-143	Sequence 143, App
19	199	2.7	432	US-08-933-750C-47	Sequence 47, Appl
20	199	2.7	432	US-09-234-613-47	Sequence 47, Appl
21	198	2.6	297	US-08-580-545B-6	Sequence 6, Appl
22	198	2.6	297	US-09-262-653A-6	Sequence 6, Appl
23	196.5	2.6	1315	US-08-899-595-3	Sequence 3, Appl
24	194.5	2.6	1248	US-09-080-897-2	Sequence 2, Appl
25	194.5	2.6	1248	US-09-323-735-2	Sequence 2, Appl
26	189.5	2.5	330	US-08-642-255-32	Sequence 32, Appl
27	189.5	2.5	408	US-07-609-716-65	Sequence 65, Appl

28	189.5	2.5	408	US-08-475-411A-65	Sequence 65, Appl
29	189.5	2.5	408	US-08-478-029A-65	Sequence 65, Appl
30	188.5	2.5	1319	US-08-290-731C-2	Sequence 2, Appl
31	188.5	2.5	1336	US-08-290-731C-6	Sequence 6, Appl
32	187.5	2.5	2414	US-08-227-536-2	Sequence 2, Appl
33	187.5	2.5	2414	PCT-US95-04682-2	Sequence 2, Appl
34	187	2.5	1312	US-09-041-886-19	Sequence 19, Appl
35	186.5	2.5	214	US-08-217-337-4	Sequence 4, Appl
36	178.5	2.4	1333	US-09-356-952-2	Sequence 2, Appl
37	175	2.3	1162	US-08-728-333A-2	Sequence 2, Appl
38	173	2.3	826	US-07-638-431-2	Sequence 2, Appl
39	173	2.3	826	PCT-US92-00018-2	Sequence 2, Appl
40	172.5	2.3	1290	US-09-150-460B-6	Sequence 6, Appl
41	170.5	2.3	614	PCT-US95-03236-21	Sequence 21, Appl
42	170.5	2.3	1291	US-09-150-460B-10	Sequence 10, Appl
43	170.5	2.3	1291	US-09-220-641-5	Sequence 5, Appl
44	170	2.3	1297	US-08-290-731C-4	Sequence 4, Appl
45	168	2.2	398	US-09-303-064-54	Sequence 54, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-869-674-2
; Sequence 2, Application US/08869674
; Patent No. 5866365
GENERAL INFORMATION:
APPLICANT: Lonetto, Michael A.
APPLICANT: Rosenberg, Martin
TITLE OF INVENTION: NOVEL COMPOUNDS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NJ
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,674
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-869-674-2
Query Match 3.7%; Score 277.5; DB 2; Length 232;
Best Local Similarity 32.5%; Pred. No. 3.6e-14;
Matches 75; Conservative 47; Mismatches 92; Indels 17; Gaps 8;
```

Db 6 TVLKNHFALEFADKNLLETAFTHTSYANEHRLKISHNERLEFLGDAVILLOLISEYLYKK 65

Qy 1166 FPDHHEGHLTLRSSLVNNRTOAKVAEELGMOEVAITNDKTRPVGLRTK--TLADLLES 1223

Db 66 YKKRKEGDLISKRAMIVRESLAFARDCQFQDF--IKLKGEEKSGGRRNRTIILGADAEFA 124

Qy 1224 FTAALYTDKLEEVHTFMNVCFPPRLK--EFLINDMNDPKSOLQOCCILTRTEGKEPDI 1281

Db 125 FIGALLLDKDVAKKEEIVQVMIKPYEAGEFEMITDY---KTHLQE--LLQVNG--DV 175

Qy 1282 PL-YKTLQVGPSHARTYVAVYFKGERIGCGKSPISQOAEKAMDALEK 1331

Db 176 AIRQVISETPAHDKVDFVEVLVEGKSIGOGGRSKRLAEQAKNAVYK 226

RESULT 2

US-09-213-010-2

; Sequence 2, Application US/09213010

; Patent No. 6251630

; GENERAL INFORMATION:

; APPLICANT: Lonetto, Michael A.

; APPLICANT: Rosenberg, Martin

; TITLE OF INVENTION: NOVEL COMPOUNDS

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert Price & Rhoads

; STREET: 997 Lenox Drive, Building 3, Suite 210

; CITY: Lawrenceville

; STATE: NJ

; COUNTRY: USA

; ZIP: 08543

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/213.010

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/869,674

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Bloom, Allen

; REGISTRATION NUMBER: 29,135

; REFERENCE/DOCKET NUMBER: GMI0013

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 609-520-3214

; TELEFAX: 609-520-3259

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 232 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-213-010-2

Query Match 3.7%; Score 277.5; DB 4; Length 232;

Best Local Similarity 32.5%; Pred. No. 3.6e-14;

Matches 75; Conservative 47; Mismatches 92; Indels 17; Gaps 8;

Qy 1108 TTEEEALIGVITFHVRLAFAFLTKTVGFNH--LTLGHNORMEFLDSDIMQVATFYTLFIH 1165

Db 6 TVLKNHFALEFADKNLLETAFTHTSYANEHRLKISHNERLEFLGDAVILLOLISEYLYKK 65

Qy 1166 FPDHHEGHLTLRSSLVNNRTOAKVAEELGMOEVAITNDKTRPVGLRTK--TLADLLES 1223

Db 66 YKKRKEGDLISKRAMIVRESLAFARDCQFQDF--IKLKGEEKSGGRRNRTIILGADAEFA 124

Qy 1224 FTAALYTDKLEEVHTFMNVCFPPRLK--EFLINDMNDPKSOLQOCCILTRTEGKEPDI 1281

Db 125 FIGALLLDKDVAKKEEIVQVMIKPYEAGEFEMITDY---KTHLQE--LLQVNG--DV 175

Qy 1282 PL-YKTLQVGPSHARTYVAVYFKGERIGCGKSPISQOAEKAMDALEK 1331

Db 176 AIRQVISETPAHDKVDFVEVLVEGKSIGOGGRSKRLAEQAKNAVYK 226

RESULT 3

US-08-965-087-2

; Sequence 2, Application US/08965087

; Patent No. 6346393

; GENERAL INFORMATION:

; APPLICANT: Fedon, Jason C.

; APPLICANT: Lonetto, Michael A.

; APPLICANT: Marra, Andrea

; APPLICANT: Palmer, Leslie M.

; APPLICANT: Rosenberg, Martin

; APPLICANT: Warren, Richard L.

; TITLE OF INVENTION: NOVEL RNCS

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert Price & Rhoads

; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

; CITY: Philadelphia

; STATE: PA

; COUNTRY: US

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/965,087

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Dickinson, Todd O

; REGISTRATION NUMBER: 28,354

; REFERENCE/DOCKET NUMBER: GMI0116

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-994-2252

; TELEFAX: 215-994-2222

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 243 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-965-087-2

Query Match 3.6%; Score 267; DB 4; Length 243;

Best Local Similarity 30.5%; Pred. No. 2.6e-13;

Matches 75; Conservative 55; Mismatches 92; Indels 24; Gaps 9;

Qy 1092 NTDRQLIETSPVLOKLTTEFEBAIGVITFHVRLAFAFLTV--GNNHLTLGNORMEFL 1149

Db 11 NREKRRDPT-----KMT-----LGFYQNDLVQOAFSSHSFINDNMNMLDNERLEFL 61

Qy 1150 GDSIMQVATFYTLFIHPRDHEGHLTLRSSLVNNRTOAKVAEELGMOEVAITNDKTRPV 1209

Db 62 GDAVELEIVSYLFDKHPNLPBGMLTKMRATVCEPSLVTFANKIGLNEKITLG-KGEER 120

Qy 1210 VGLRTK--TLADLLESFIALYTDKLEEVHTFMNVCFPPRLKEFLINDMND--DKRSQ 1265

Db 121 TGGTRSLISDAFEALIGALYLDGGLDIYMKFAEKYIFPHVQ---NELIGVVDKRTQF 177

Qy 1266 QOCCILTRTEGKEPDIPLYTLQVGPSHARTYVAVYFKGERIGCGKSPISQOAEKMA 1325

Db 178 OEYV---HQONKGDV-TYNLIKEGPAHHRLFTSEVILLQGBAIAEGKGTKESEORAA 232  
QY 1326 MDALER 1331  
Db 233 ESAYKO 238

RESULT 4  
5202236-3

; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
; PROTEIN

; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/528,762

; FILING DATE: 25-MAY-1990

; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987

; APPLICATION NUMBER: 933,945

; FILING DATE: 24-NOV-1986

; APPLICATION NUMBER: 650,128

; FILING DATE: 13-SEP-1984

; SEQ ID NO:3:

; LENGTH: 334

5202236-3

Query Match 3.0%; Score 226; DB 6; Length 334;  
Best Local Similarity 31.8%; Pred. No. 7e-10;  
Matches 61; Conservative 21; Mismatches 70; Indels 40; Gaps 10;

QY 55 PRPAPRTTNSNPAPFLPRPDVFPFRPMPBSAGRLPCGIRPFPHQMRHFPVP 114

Db 5 PRPAPRAPAPALAPAPNPV--PQSPSPSPPTP--PPTPPSPAP-----BSP 53

QY 115 PCPPAPPPMPCNNPVPAGPQGTFP--FMPPPSMHPPPVMPQVNY----- 166

Db 54 PRSPRPPTSP--PRSPRPAPRPSPFPPTKAKTYAKKTNTSTYAKKSYPTTYKA 112

QY 167 --QYPPGY-SHNFP-----PSFNSFQNNPSSFLPSANSSS-----PHFRHLPY 210

Db 113 KSPPTTYAKKSPPTTYAKKPTTKIKPTYPSTYKAKPTNPSTYKAKPSYPTTYAKPSY 172

QY 211 PLKAPSERRSP 222

Db 173 P-----PTYKAKP 180

RESULT 5  
US-09-095-443-2

; Sequence 2, Application US/09095443  
; Patent No. 6342593

; GENERAL INFORMATION:

; APPLICANT: PLOWMAN, Gregory

; APPLICANT: Peles, Eliot

; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; CITY: Suite 4700

; STATE: Los Angeles

; COUNTRY: U.S.A.

; ZIP: 90071-2066

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: FastSEQ for Windows 2.0

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: us/09/095,443  
; FILING DATE: Herewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/049,477

; FILING DATE: June 12, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Warburg, Richard J.

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 235/055

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELETYPE: 67-3510

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1274 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-095-443-2

Query Match 3.0%; Score 226; DB 4; Length 1274;  
Best Local Similarity 29.8%; Pred. No. 4.8e-09;  
Matches 115; Conservative 31; Mismatches 144; Indels 96; Gaps 25;

QY 8 HRMSFHPGRCGRGKGHAR-----SAPSRQNLRLHQQP--PVQYQY--EPSPAPS 60

Db 574 HNFSSGIPTGFPAPRIGPQRPQRPHPQSQAFGPQ-----PQOPRLIQNHLEPPQAPG 628

QY 61 TTFSSPAPNLPFRPDVFPFRPMPBSAGRLPCGIRPFPHQMRHFPVP--PCFP 119

Db 629 LLPQSPPT--YAPQGVGLQRP--PLHTQ--LTPGPAQDLPRASGALRPPSGPQRP 683

QY 120 MPPEW--PCPNPNPV--PGAP-----GQGT--FFPMPPSPMPRPVMPQVNY 166

Db 684 HPPLAYGAPATRTMGPAQLITRGSSAGQSPPHILVSPRA--PAPGQVPPR----- 738

QY 167 QYPPGYSHNFP-----PSFNSFQNNPSSFLPSANSSSPPHFR-----LPPYLPK 214

Db 739 --PPAAE-----PPCLR-----RGAAAADLLSSPSQHGQSPGCGOPLLPQTVDA 786

QY 215 APSESRPERLKHVDHNRHDSHGGRHRSIDLR---ERGRSDRRDSDRSYD- 270

Db 787 AEGRRQALRLERDPYEH-----PERLROLQQLQELAEFGQLGDVGLDTVRELQDA 839

QY 271 -----RGRTPSRHSYERSRERERHNRHD-----NRRSPSLERSYKKEYSKSGSYG 319

Db 840 QEHDAKRGSIATACYSL-----KNRHQDVMYDYSNR---VYLRSGKDYIASCVEG 889

QY 320 LSVVPEPAGCT--PELPEIINKTDSW 344

Db 890 LSPYCPPLVATQAPLPGFA---ADFW 912

RESULT 6  
5202236-37

; Patent No. 5202236

; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID

; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE

; PROTEIN

; NUMBER OF SEQUENCES: 39

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/528,762

; FILING DATE: 25-MAY-1990

; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987

; APPLICATION NUMBER: 933,945

; FILING DATE: 24-NOV-1986

; APPLICATION NUMBER: 650,128



TELEFAX: 608-251-9166  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 306 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-217-327-6

Query Match 2.9%; Score 218; DB 1; Length 306;  
 Best Local Similarity 27.6%; Pred. No. 2.6e-09;  
 Matches 68; Conservative 14; Mismatches 96; Indels 68; Gaps 10;

QY 47 PPVYOYE-----PPSAPSTTFNSPAPNLPRLPRDPVPRPPMPSPAGPL----- 93  
 Db 53 PPRVYHESPPRRKHSPPRRPVYKYSPPRRMHSPPRRVNFESPPRRKHSPPRRPVYK 112  
 QY 94 ---PSPTRPPRPNHQRNRPVPC-----PSPMPRPPCP-----NNPPV----- 133  
 Db 113 YKSPRRKHSAPVNHKYSPPRRPVYKYSPPRRKHSAPVNHKYSPPRRKHFPA 172  
 QY 134 -----GAPGOSTFRPMMPSP-----HPPRRVPMQOV-NYQY---PQYSH 174  
 Db 173 PBNHYKYSKYSPPRRPVYKYSPPRRPVYKYSPPRRKHSAPVNHKYSPPRRPVY 232  
 QY 175 HNPFPFSFNSPFPSSFLPSANSSSPHFRHLPPVLP-----KAPSESRSP 222  
 Db 233 YKSPRP-----EHSPPRRPVYKYSPPRRMHSPPRRPVYKYSPPRRMHSPPRPVSP 288  
 QY 223 ERLKHY 228  
 Db 289 PPKKH 294

RESULT 9  
 US-08-574-959A-9  
 Sequence 9, Application US/08574959A  
 Patent No. 5962224  
 GENERAL INFORMATION:  
 APPLICANT: Jaekyoon Shin, Insil Jounq, Ratna K. Vladlamudi  
 TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
 TITLE OF INVENTION: AND USES THEREFOR  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 60 State Street, Suite 510  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/574,959A  
 FILING DATE: 19-DEC-95  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mandragouras, Amy E.  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: DFN-008  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)227-7400  
 TELEFAX: (617)227-5941  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 905 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-574-959A-9

Query Match 2.9%; Score 216; DB 2; Length 905;  
 Best Local Similarity 21.2%; Pred. No. 1.8e-08;  
 Matches 125; Conservative 58; Mismatches 212; Indels 196; Gaps 29;

QY 6 TCHRMSPH-----GRCGRCGRGCGHAGRSAPSFSPON-LRLHPOOPVQYQYRPP 56  
 Db 394 TCALAL-HRVPRLQPMGTCTPAVPLRLRHRSGRHSLRLAPCQMACPCQACPC 452  
 QY 57 SA-----PSTFNSPA-----PNFLPP-----PDPVP 80  
 Db 453 SAGPMPSAGVPSEPTSTANLGLLSRVSVCPRLLPGPNHRSAGSNEDPILASGTP 512  
 QY 81 FPPMPSPNQ---GRLPSPCTIRPPFRPNHQRNRP-----VPPCFPPMPR 123  
 Db 513 -PRTTPRDEFGGRVP---RPAFVHYKDEASDVEITSLSDSDSVIVVPGRLPLPP 567  
 QY 124 MPCRNRPVYGARPGQSTFRMP-----PSMPHPPRPVPMQOVNYQY 168  
 Db 568 PPSGATPP-PIAPTP---PTASPPVPAKEEDELPAAPGLRPPPPPPPPVPGPV----- 619  
 QY 169 PPGYSHNFPSPFSFNSPFLPSA-----NNSSPFRHLPPVLPKAPSERR 220  
 Db 620 -----XLPPLQVP-EGTPGGGPPALEEDLTVININSDE-----EEEE 658  
 QY 221 SPERLKHVDHNRHDSHGCRHSRLDRRGRS---PDRRQDSRYRQVDRGRTPSR 277  
 Db 659 EGDEEEDEEEDEEEDEEEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 711  
 QY 278 HRSYERSRERERHNRHNRSPSLERSYKKEKYSGRGSLVYPVAGCTPLPGEI 337  
 Db 712 ---EEEGELE-EEDEEDEEEDEEEDEVEDELEFGTAG-----GEVEGAPPPPLPAL 761  
 QY 338 IKNTDSMAPPLEIVNHRSSREKKRAMEEKDRSDNOSGKDKNYISIKKEPEETMP 397  
 Db 762 -----PPPSPPKQPEPEPEGLLLEVEEPGTE-EEEGADAPPTL-----APEALP 807  
 QY 398 DKNEEEELRLPVMIRCHSENYSSDPMDQVSTVGTSLRLDLYDK-----FE 449  
 Db 808 SOGEVRE-----GESPAAGPP-OLVEEESXPTTLE 841  
 QY 450 EELGSHOEKAKAAPRPPKTKLDELSSSESRC-ESDEDSCTSSSD 498  
 Db 842 EETEDGSDKVQ-----PPPTPAEEMETETEALQEKQDDPAAMLAD 886

RESULT 10  
 US-09-357-014-9  
 Sequence 9, Application US/09357014  
 Patent No. 6291645  
 GENERAL INFORMATION:  
 APPLICANT: Jaekyoon Shin, Insil Jounq, Ratna K. Vladlamudi  
 TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
 TITLE OF INVENTION: AND USES THEREFOR  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 60 State Street, Suite 510  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/357,014  
 FILING DATE: 19-Jul-1999  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/574,959

```

; FILING DATE: <UNKNOWN>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-014-9

Query Match      2.9%; Score 216; DB 4; Length 905;
Best Local Similarity 21.2%; Pred. No. 1.8e-08;
Matches 125; Conservative 58; Mismatches 212; Indels 196; Gaps 29;

QY 6 TCHMSFHP-----GRCPRGRGNGGARSPARFRPQN-LRLHPQOQPVYQYQYEP 56
DB 394 TCAALT-HRVPRLQPMGTCPTPARVPLLRHPSGPHRSILRAPCFQMAPCQOQAPCP 452
QY 57 SA-----PSTTFSNSPA-----PNFLPPR-----PDFVP 80
DB 453 SAGBPMGAGVPSPMTSTANLGLLSRPSVCPRLRPGENHRAGSNEDPILASGTP 512
QY 81 FPPMPPSAQ--GLRPPCIRPPFPNHNQMRPP-----VPPCFPPMP 123
DB 513 -PPTIPDETFTGGVVP---RPAFVHYDKEAASDVETLSLSDSDSVIVYVSGLPPLPP 567
QY 124 MCPNNRPVPGAPPGGCTPFPMMP-----PSMHPPPPVMPQOQVNYQY 168
DB 568 PPSGATPP-PIAPTGP---PTASPPVPAKKEPEBELPAABGRLPPPPPPVPGV----- 619
QY 169 PGYSHHNPPPSFNSFQNNPSSFLPSA-----NNSSPHFHRLPPYLPKAPSER 220
DB 620 -----XLRPPQLVP-EGTPGGGPPALBEDLTVININSSDE-----EEEE 658
QY 721 SPEPLKHYDDHNRHDSHGGRHRSIDRRERGRS---PDRRQDSRYSIDYDKRTSR 277
DB 659 EGEDEEEDEEEDEEEDEEEDEFEDEDEDEEYEEDEEEDEEEDE----- 711
QY 728 HRSYERERERERHRNRHRSRPSLSRYKKYKRGSGYGLSVVPAGCTPELPGEI 337
DB 712 -----EEEGELE-EEDEEDEEEDEELEEVEDLEFGTAG-----GEVEGAPPPPTLPAL 761
QY 338 IKNTDSWAPPLEIVNHRSPSRKKRARWEERKDRWSDNOSGDKNYTSIKEKEPETMP 397
DB 762 -----PPESPCKVQPEPEPGLLVEVEPGTE-EEGADTAFTL-----APEALP 807
QY 398 DKNEEEELLPVWIRCHSENYSDDPMQVGDSTVGTSLRLDYDK-----FE 449
DB 808 SQGEVERE-----GESPAAGPPP-QELVEEPEXPPTLLE 841
QY 450 EELGSRQEKAKAAPPEPKTKLDEDLSSSESEC--ESDEDSTCSSSSD 498
DB 842 EETEDGSDKVQ-----PPETPAEEMETETEAALQEKEDDTAAMLAD 886

RESULT 11
US-08-574-959A-7
; Sequence 7, Application US/08574959A
; Patent No. 5962224
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Jung, Ratna K. Vadlamudi
; APPLICANT: and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/574, 959A
; FILING DATE: 19-DEC-95
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-574-959A-7

Query Match      2.9%; Score 216; DB 2; Length 1135;
Best Local Similarity 21.2%; Pred. No. 2.5e-08;
Matches 125; Conservative 58; Mismatches 212; Indels 196; Gaps 29;

QY 6 TCHMSFHP-----GRCPRGRGNGGARSPARFRPQN-LRLHPQOQPVYQYQYEP 56
DB 624 TCAALT-HRVPRLQPMGTCPTPARVPLLRHPSGPHRSILRAPCFQMAPCQOQAPCP 682
QY 57 SA-----PSTTFSNSPA-----PNFLPPR-----PDFVP 80
DB 683 SAGBPMGAGVPSPMTSTANLGLLSRPSVCPRLRPGENHRAGSNEDPILASGTP 742
QY 81 FPPMPPSAQ--GLRPPCIRPPFPNHNQMRPP-----VPPCFPPMP 123
DB 743 -PPTIPDETFTGGVVP---RPAFVHYDKEAASDVETLSLSDSDSVIVYVSGLPPLPP 797
QY 124 MCPNNRPVPGAPPGGCTPFPMMP-----PSMHPPPPVMPQOQVNYQY 168
DB 798 PPSGATPP-PIAPTGP---PTASPPVPAKKEPEBELPAABGRLPPPPPPVPGV----- 849
QY 169 PGYSHHNPPPSFNSFQNNPSSFLPSA-----NNSSPHFHRLPPYLPKAPSER 220
DB 850 -----XLRPPQLVP-EGTPGGGPPALBEDLTVININSSDE-----EEEE 888
QY 221 SPEPLKHYDDHNRHDSHGGRHRSIDRRERGRS---PDRRQDSRYSIDYDKRTSR 277
DB 889 EGEDEEEDEEEDEEEDEEEDFEDEDEDEEYEEDEEEDEEEDE----- 941
QY 278 HRSYERERERERHRNRHRSRPSLSRYKKYKRGSGYGLSVVPAGCTPELPGEI 337
DB 942 -----EEEGELE-EEDEEDEEEDEELEEVEDLEFGTAG-----GEVEGAPPPPTLPAL 991
QY 338 IKNTDSWAPPLEIVNHRSPSRKKRARWEERKDRWSDNOSGDKNYTSIKEKEPETMP 397
DB 992 -----PPESPCKVQPEPEPGLLVEVEPGTE-EEGADTAFTL-----APEALP 1037
QY 398 DKNEEEELLPVWIRCHSENYSDDPMQVGDSTVGTSLRLDYDK-----FE 449
DB 1038 SQGEVERE-----GESPAAGPPP-QELVEEPEXPPTLLE 1071
QY 450 EELGSRQEKAKAAPPEPKTKLDEDLSSSESEC--ESDEDSTCSSSSD 498
DB 1072 EETEDGSDKVQ-----PPETPAEEMETETEAALQEKEDDTAAMLAD 1116
```







Search completed: October 9, 2002, 18:27:44  
Job time : 46 secs

2



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 18:24:08 : Search time 200 Seconds  
(without alignments)  
1900.652 Million cell updates/sec

Title: US-09-900-425a-2  
Perfect score: 7500  
Sequence: 1 MNOGNTCHRMSPHPCGRCPR.....MRWERHQRREDEDEDIKK 1374

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1006125 seqs, 276659714 residues

Total number of hits satisfying chosen parameters: 1006125

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PC1\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4071	54.3	769	5	US-09-629-469A-10949
2	2847	38.0	541	6	US-10-205-331-116
3	1961	26.1	378	5	US-09-764-882A-579
4	1961	26.1	378	6	US-10-103-313-471
5	1956	26.1	378	5	US-09-764-882A-377
6	1320	17.6	263	6	US-10-103-313-307
7	496	6.6	115	5	US-09-547-599C-3281
8	315	4.2	249	5	US-09-791-537-130915
9	309	4.1	266	5	US-09-791-537-62108
10	305	4.1	266	5	US-09-791-537-104336
11	286	3.8	480	6	US-10-216-583-1603
12	273	3.6	827	5	US-09-935-625-17077
13	273	3.6	847	5	US-09-935-625-17076
14	273	3.6	859	5	US-09-935-625-17480
15	273	3.6	879	5	US-09-935-625-17479
16	273	3.6	891	5	US-09-935-625-17075
17	273	3.6	903	5	US-09-935-625-17478
18	267	3.6	224	5	US-09-791-537-55409
19	265.5	3.5	1002	5	US-09-791-537-143780
20	263.5	3.5	1130	6	US-10-179-131-9979
21	262	3.5	226	7	US-60-402-347-1
22	259.5	3.5	1589	5	US-09-935-625-9406
23	259.5	3.5	1589	5	US-09-935-625-26246
24	259.5	3.5	1615	5	US-09-935-625-9405
25	259.5	3.5	1615	5	US-09-935-625-26245
26	259.5	3.5	1615	5	US-09-935-625-26245

27	259.5	3.5	1649	5	US-09-935-625-9404	Sequence 9404, Ap
28	259.5	3.5	1649	5	US-09-935-625-26244	Sequence 26244, A
29	258	3.4	231	5	US-09-791-537-38286	Sequence 38286, A
30	258	3.4	598	1	PCT-US02-25485-2790	Sequence 2790, Ap
31	257	3.4	1250	1	PCT-US02-16876-9	Sequence 9, App11
32	256	3.4	246	5	US-09-791-537-130365	Sequence 130365,
33	254.5	3.4	648	6	US-10-179-131-5134	Sequence 5134, Ap
34	254	3.4	2249	5	US-09-866-557A-4	Sequence 4, App11
35	254	3.4	2249	5	US-09-858-862-4	Sequence 4, App11
36	254	3.4	2249	6	US-10-055-797-4	Sequence 130916,
37	251.5	3.4	239	5	US-09-791-537-130916	Sequence 80004, A
38	251.5	3.4	240	5	US-09-791-537-80004	Sequence 900, App
39	250.5	3.3	592	5	US-09-935-625-900	Sequence 5343, Ap
40	250.5	3.3	592	5	US-09-935-625-5343	Sequence 8099, Ap
41	250.5	3.3	592	5	US-09-935-625-8099	Sequence 25320, A
42	250.5	3.3	592	5	US-09-935-625-25320	Sequence 899, App
43	250.5	3.3	597	5	US-09-935-625-899	Sequence 5342, Ap
44	250.5	3.3	597	5	US-09-935-625-5342	Sequence 8098, Ap
45	250.5	3.3	597	5	US-09-935-625-8098	

## ALIGNMENTS

```
RESULT 1
US-09-629-469A-10949
: Sequence 10949, Application US/09629469A
:
: GENERAL INFORMATION:
: APPLICANT: OTA, TOSHIO
: APPLICANT: ISOGAI, TAKAO
: APPLICANT: NISHIKAWA, TETSUO
: APPLICANT: HAYASHI, KOJI
: APPLICANT: SATTO, KAOBU
: APPLICANT: YAMAMOTO, JUNICHI
: APPLICANT: ISHII, SHIZUKO
: APPLICANT: SUGIYAMA, TOMOYASU
: APPLICANT: WAKAMATSU, AI
: APPLICANT: NAGAI, KEIICHI
: APPLICANT: OTSUKI, TETSUJI
: TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE
: FILE REFERENCE: 084335/0123
: CURRENT FILING DATE: 2000-07-28
: PRIOR APPLICATION NUMBER: JP 1999-248036
: PRIOR FILING DATE: 1999-07-29
: PRIOR APPLICATION NUMBER: JP 1999-300253
: PRIOR FILING DATE: 1999-08-27
: PRIOR APPLICATION NUMBER: JP 2000-118776
: PRIOR FILING DATE: 2000-01-11
: PRIOR APPLICATION NUMBER: JP 2000-183767
: PRIOR FILING DATE: 2000-05-02
: PRIOR APPLICATION NUMBER: JP 2000-241899
: PRIOR FILING DATE: 2000-06-09
: PRIOR APPLICATION NUMBER: 60/159,590
: PRIOR FILING DATE: 1999-10-18
: PRIOR APPLICATION NUMBER: 60/183,322
: PRIOR FILING DATE: 2000-02-17
: NUMBER OF SEQ ID NOS: 19025
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 10949
:
: TYPE: PRT
: ORGANISM: Homo sapiens
:
: US-09-629-469A-10949
:
: Query Match 54.3% Score 4071; DB 5; Length 769;
: Best Local Similarity 99.5%; Pred. No. 4e-185;
: Matches 765; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 606 MEAHPPLNIPICRYIRINIDYTHIFTEAMPENPCVKGLEFLFRLDILEYDNNLK 665
DB 1 MEAHPPLNIPICRYIRINIDYTHIFTEAMPENPCVKGLEFLFRLDILEYDNNLK 60
```

```

QY 666 GFLFEDSPCCRFHFHMFVNFVFLPDGGEVLSMHQIILLYLLRCSKALVPEEETIANLQW 725
    |||||||
Db 61 GFLFEDSPCCRFHFHMFVNFVFLPDGGEVLSMHQIILLYLLRCSKALVPEEETIANLQW 120
QY 726 EELBQKTAEECKGMITVNPGRKSSVRIQDLDRQENPDVITPTIYHSGIRPAQLSYA 785
    |||||||
Db 121 EELBQKTAEECKGMITVNPGRKSSVRIQDLDRQENPDVITPTIYHSGIRPAQLSYA 180
QY 786 GPOYOKLWKSYYVLRHLLANSPPKQOTDKOKLAOREALQKIROKNTMREVEVTEISSQ 845
    |||||||
Db 181 GPOYOKLWKSYYVLRHLLANSPPKQOTDKOKLAOREALQKIROKNTMREVEVTEISSQ 240
QY 846 GFWKTGIRSDVCQHAMLPLVLTNHIRYHQCIMHLDKLGTYTFODRCILLQAMTHPSHHLN 905
    |||||||
Db 241 GFWKTGIRSDVCQHAMLPLVLTNHIRYHQCIMHLDKLGTYTFODRCILLQAMTHPSHHLN 300
QY 906 FGAMNDHANSLNSGIRPKYGDGRKHVHHMRRKKGINTLTINMSRLGODDPTPSRIHN 965
    |||||||
Db 301 FGAMNDHANSLNSGIRPKYGDGRKHVHHMRRKKGINTLTINMSRLGODDPTPSRIHN 360
QY 966 ERLLEFGDAVVEFLTSVHLHYLFLPSLEEGGLATYRTAIVONQHLAMAKLELDFMFLYA 1025
    |||||||
Db 361 ERLLEFGDAVVEFLTSVHLHYLFLPSLEEGGLATYRTAIVONQHLAMAKLELDFMFLYA 420
QY 1026 HGPDLCRESDLRHMANCFEALIGAIVLEGSLEAKOLFGRLLFNDPDLREVWNLPLAP 1085
    |||||||
Db 421 HGPDLCRESDLRHMANCFEALIGAIVLEGSLEAKOLFGRLLFNDPDLREVWNLPLAP 480
QY 1086 ILOLEPNDRQOLIETSPVLOKLTFFEEAIGVIFTHVRLARAFTLRVGFNHLTIGHNOR 1145
    |||||||
Db 481 ILOLEPNDRQOLIETSPVLOKLTFFEEAIGVIFTHVRLARAFTLRVGFNHLTIGHNOR 540
QY 1146 MEFLGDSIMQVATEYLFTHPPDHHEGHLTLRSSLVNNRTQAKVAEELGMOEYAITNDK 1205
    |||||||
Db 541 MEFLGDSIMQVATEYLFTHPPDHHEGHLTLRSSLVNNRTQAKVAEELGMOEYAITNDK 600
QY -1206 TRPVGLRKTTLADLESTIALYTDKOLEVYHTPMNVCFFPRLEKEFTLNDMDPKSOL 1265
    |||||||
Db 601 TRPVGLRKTTLADLESTIALYTDKOLEVYHTPMNVCFFPRLEKEFTLNDMDPKSOL 660
QY 1266 QOCCLTLRTEGKEPDIPYKTLQTVGSPSHARTYVAAYFKGERIGCGKSPISQAEWGAA 1325
    |||||||
Db 661 QOCCLTLRTEGKEPDIPYKTLQTVGSPSHARTYVAAYFKGERIGCGKSPISQAEWGAA 720
QY 1326 MDALERYNPQMAHOKRFTGRKYRQELKEMWREHOREDEDEDIK 1374
    |||||||
Db 721 MDALERYNPQMAHOKRFTGRKYRQELKEMWREHOREDEDEDIK 769

RESULT 2
US-10-205-331-116
; Sequence 116, Application US/10205331
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pincock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WI-A-018199
; CURRENT APPLICATION NUMBER: US/10/205,331
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 116
; LENGTH: 541
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Putative ribonuclease III
US-10-205-331-116

```

```

Query Match          38.0%; Score 2847; DB 6; Length 541;
Best Local Similarity 99.3%; Pred. No. 2,8e-127;
Matches 537; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 834 MRREVEVTEISSQGFWMKTGIRSDVCQHAMLPLVLTNHIRYHQCIMHLDKLGTYTFODRCIL 893
    |||||||
Db 1 MRREVEVTEISSQGFWMKTGIRSDVCQHAMLPLVLTNHIRYHQCIMHLDKLGTYTFODRCIL 60
QY 894 QLAMTHPSHHLNFGAMPDHNARNLSNCGITROPYGGRKVHHMRRKKGINTLTINMSRLG 953
    |||||||
Db 61 QLAMTHPSHHLNFGAMPDHNARNLSNCGITROPYGGRKVHHMRRKKGINTLTINMSRLG 120
QY 954 QDDPTPSRIHNRELEFLGDAVVEFLTSVHLHYLFLPSLEEGGLATYRTAIVONQHLAMLA 1013
    |||||||
Db 121 QDDPTPSRIHNRELEFLGDAVVEFLTSVHLHYLFLPSLEEGGLATYRTAIVONQHLAMLA 180
QY 1014 KLELDFMFLYAHGPDLCRESDLRHMANCFEALIGAIVLEGSLEAKOLFGRLLFNDPD 1073
    |||||||
Db 181 KLELDFMFLYAHGPDLCRESDLRHMANCFEALIGAIVLEGSLEAKOLFGRLLFNDPD 240
QY 1074 LREWNLPLAPLOLEPNDRQOLIETSPVLOKLTFFEEAIGVIFTHVRLARAFTLRV 1133
    |||||||
Db 241 LREWNLPLAPLOLEPNDRQOLIETSPVLOKLTFFEEAIGVIFTHVRLARAFTLRV 300
QY 1134 GFNHLTIGHNORMEFLGDSIMQVATEYLFTHPPDHHEGHLTLRSSLVNNRTQAKVAE 1193
    |||||||
Db 301 GFNHLTIGHNORMEFLGDSIMQVATEYLFTHPPDHHEGHLTLRSSLVNNRTQAKVAE 360
QY 1194 LGMOEYAITNDKTRPVGLRKTTLADLESTIALYTDKOLEVYHTPMNVCFFPRLEKEFT 1253
    |||||||
Db 361 LGMOEYAITNDKTRPVGLRKTTLADLESTIALYTDKOLEVYHTPMNVCFFPRLEKEFT 420
QY 1254 LNDMDPKSOLQOCCLTLRTEGKEPDIPYKTLQTVGSPSHARTYVAAYFKGERIGCGK 1313
    |||||||
Db 421 LNDMDPKSOLQOCCLTLRTEGKEPDIPYKTLQTVGSPSHARTYVAAYFKGERIGCGK 480
QY 1314 GPSIQAEWGAAAMDALERYNPQMAHOKRFTGRKYRQELKEMWREHOREDEDEDIK 1373
    |||||||
Db 481 GPSIQAEWGAAAMDALERYNPQMAHOKRFTGRKYRQELKEMWREHOREDEDEDIK 540
QY 1374 K 1374
    |||||||
Db 541 K 541

RESULT 3
US-09-764-882A-579
; Sequence 579, Application US/09764882A
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT215
; CURRENT APPLICATION NUMBER: US/09/764,882A
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14

```

PRIOR APPLICATION NUMBER: 60/225,757  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/226,868  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/216,647  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,267  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/216,880  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,270  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/251,869  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/235,834  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/234,274  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/234,223  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/228,924  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/224,518  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,369  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/224,519  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,964  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/241,809  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/249,299  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/236,327  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/241,785  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/244,617  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/225,268  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802

PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232,400  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/231,242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,080  
PRIOR FILING DATE: 2000-09-08

```

; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,065
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,398
```

```

Query Match          26.1%; Score 1961; DB 5; Length 378;
Best Local Similarity 98.9%; Pred. No. 1.8e-85;
Matches 374; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

QY 997 ATYTAIVONQHMLAMAKKLEDPFMYLHAGPDLRESDLRHMANCFEALIGAVYLEGS 1056
    |||||||
DB 1 ATYTAIVONQHMLAMAKKLEDPFMYLHAGPDLRESDLRHMANCFEALIGAVYLEGS 60
QY 1057 LEEAKOLFGRLLFNDPDLREVMYLPPLHPLQLOEPNTDROLITSPVLQKLFEEAIGV 1116
    |||||||
DB 61 LEEAKOLFGRLLFNDPDLREVMYLPPLHPLQLOEPNTDROLITSPVLQKLFEEAIGV 120
QY 1117 IFTHVRLARAFTLRTVGFNHLTLGHNORMEFLGDSIMQVATEYLFIFHPDHHEGHLTL 1176
    |||||||
DB 121 IFTHVRLARAFTLRTVGFNHLTLGHNORMEFLGDSIMQVATEYLFIFHPDHHEGHLTL 180
QY 1177 LRSSLVNNRTQAKVAEELGMEYAITNDKTRPVGLRTKTLADLLESFIALYTDKLEY 1236
    |||||||
DB 181 LRSSLVNNRTQAKVAEELGMEYAITNDKTRPVGLRTKTLADLLESFIALYTDKLEY 240
QY 1237 VHTFMNVCFFPRLKEFLINQDMNDPKSLOQCCCLTLRTGKKEPDIPLYKTLQTVGSPSHAR 1296
    |||||||
DB 241 VHTFMNVCFFPRLKEFLINQDMNDPKSLOQCCCLTLRTGKKEPDIPLYKTLQTVGSPSHAR 300
QY 1297 TYTVAVYFKGERIGCGKSPISQOAEKGMADALEKYNPQMAHQKRFIGRYROELKEMR 1336
    |||||||
DB 301 TYTVAVYFKGERIGCGKSPISQOAEKGMADALEKYNPQMAHQKRFIGRYROELKEMR 360
QY 1357 WEREHOREPDETDIKK 1374
    |||||||
DB 361 WEREHOREPDETDIKK 378
```

```

RESULT 4
US-10-103-313-471
; Sequence 471, Application US/10103313
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
```

```

; PRIOR Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 471
; LENGTH: 378
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-103-313-471
```

```

Query Match          26.1%; Score 1961; DB 6; Length 378;
Best Local Similarity 98.9%; Pred. No. 1.8e-85;
Matches 374; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

QY 997 ATYTAIVONQHMLAMAKKLEDPFMYLHAGPDLRESDLRHMANCFEALIGAVYLEGS 1056
    |||||||
DB 1 ATYTAIVONQHMLAMAKKLEDPFMYLHAGPDLRESDLRHMANCFEALIGAVYLEGS 60
QY 1057 LEEAKOLFGRLLFNDPDLREVMYLPPLHPLQLOEPNTDROLITSPVLQKLFEEAIGV 1116
    |||||||
DB 61 LEEAKOLFGRLLFNDPDLREVMYLPPLHPLQLOEPNTDROLITSPVLQKLFEEAIGV 120
QY 1117 IFTHVRLARAFTLRTVGFNHLTLGHNORMEFLGDSIMQVATEYLFIFHPDHHEGHLTL 1176
    |||||||
DB 121 IFTHVRLARAFTLRTVGFNHLTLGHNORMEFLGDSIMQVATEYLFIFHPDHHEGHLTL 180
QY 1177 LRSSLVNNRTQAKVAEELGMEYAITNDKTRPVGLRTKTLADLLESFIALYTDKLEY 1236
    |||||||
DB 181 LRSSLVNNRTQAKVAEELGMEYAITNDKTRPVGLRTKTLADLLESFIALYTDKLEY 240
QY 1237 VHTFMNVCFFPRLKEFLINQDMNDPKSLOQCCCLTLRTGKKEPDIPLYKTLQTVGSPSHAR 1296
    |||||||
DB 241 VHTFMNVCFFPRLKEFLINQDMNDPKSLOQCCCLTLRTGKKEPDIPLYKTLQTVGSPSHAR 300
QY 1297 TYTVAVYFKGERIGCGKSPISQOAEKGMADALEKYNPQMAHQKRFIGRYROELKEMR 1336
    |||||||
DB 301 TYTVAVYFKGERIGCGKSPISQOAEKGMADALEKYNPQMAHQKRFIGRYROELKEMR 360
QY 1357 WEREHOREPDETDIKK 1374
    |||||||
DB 361 WEREHOREPDETDIKK 378
```

```

RESULT 5
US-09-764-882A-377
; Sequence 377, Application US/09764882A
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT215
; CURRENT APPLICATION NUMBER: US/09/764,882A
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
```



1	PRIOR APPLICATION	NUMBER: 60/240,966
2	PRIOR FILING DATE:	2000-10-20
3	PRIOR APPLICATION	NUMBER: 60/239,935
4	PRIOR FILING DATE:	2000-10-13
5	PRIOR APPLICATION	NUMBER: 60/239,937
6	PRIOR FILING DATE:	2000-10-13
7	PRIOR APPLICATION	NUMBER: 60/241,787
8	PRIOR FILING DATE:	2000-10-20
9	PRIOR APPLICATION	NUMBER: 60/246,474
10	PRIOR FILING DATE:	2000-11-08
11	PRIOR APPLICATION	NUMBER: 60/246,532
12	PRIOR FILING DATE:	2000-11-08
13	PRIOR APPLICATION	NUMBER: 60/249,216
14	PRIOR FILING DATE:	2000-11-17
15	PRIOR APPLICATION	NUMBER: 60/249,210
16	PRIOR FILING DATE:	2000-11-17
17	PRIOR APPLICATION	NUMBER: 60/226,681
18	PRIOR FILING DATE:	2000-08-22
19	PRIOR APPLICATION	NUMBER: 60/225,759
20	PRIOR FILING DATE:	2000-08-14
21	PRIOR APPLICATION	NUMBER: 60/225,213
22	PRIOR FILING DATE:	2000-08-14
23	PRIOR APPLICATION	NUMBER: 60/227,182
24	PRIOR FILING DATE:	2000-08-22
25	PRIOR APPLICATION	NUMBER: 60/225,214
26	PRIOR FILING DATE:	2000-08-14
27	PRIOR APPLICATION	NUMBER: 60/225,836
28	PRIOR FILING DATE:	2000-09-27
29	PRIOR APPLICATION	NUMBER: 60/230,438
30	PRIOR FILING DATE:	2000-09-06
31	PRIOR APPLICATION	NUMBER: 60/215,135
32	PRIOR FILING DATE:	2000-06-30
33	PRIOR APPLICATION	NUMBER: 60/225,266
34	PRIOR FILING DATE:	2000-08-14
35	PRIOR APPLICATION	NUMBER: 60/249,218
36	PRIOR FILING DATE:	2000-11-17
37	PRIOR APPLICATION	NUMBER: 60/249,208
38	PRIOR FILING DATE:	2000-11-17
39	PRIOR APPLICATION	NUMBER: 60/249,213
40	PRIOR FILING DATE:	2000-11-17
41	PRIOR APPLICATION	NUMBER: 60/249,212
42	PRIOR FILING DATE:	2000-11-17
43	PRIOR APPLICATION	NUMBER: 60/249,207
44	PRIOR FILING DATE:	2000-11-17
45	PRIOR APPLICATION	NUMBER: 60/249,245
46	PRIOR FILING DATE:	2000-11-17
47	PRIOR APPLICATION	NUMBER: 60/249,244
48	PRIOR FILING DATE:	2000-11-17
49	PRIOR APPLICATION	NUMBER: 60/249,217
50	PRIOR FILING DATE:	2000-11-17
51	PRIOR APPLICATION	NUMBER: 60/249,211
52	PRIOR FILING DATE:	2000-11-17
53	PRIOR APPLICATION	NUMBER: 60/249,215
54	PRIOR FILING DATE:	2000-11-17
55	PRIOR APPLICATION	NUMBER: 60/249,264
56	PRIOR FILING DATE:	2000-11-17
57	PRIOR APPLICATION	NUMBER: 60/249,214
58	PRIOR FILING DATE:	2000-11-17
59	PRIOR APPLICATION	NUMBER: 60/249,297
60	PRIOR FILING DATE:	2000-11-17
61	PRIOR APPLICATION	NUMBER: 60/232,400
62	PRIOR FILING DATE:	2000-09-14
63	PRIOR APPLICATION	NUMBER: 60/231,242
64	PRIOR FILING DATE:	2000-09-08
65	PRIOR APPLICATION	NUMBER: 60/232,081
66	PRIOR FILING DATE:	2000-09-08
67	PRIOR APPLICATION	NUMBER: 60/232,080
68	PRIOR FILING DATE:	2000-09-08
69	PRIOR APPLICATION	NUMBER: 60/231,414
70	PRIOR FILING DATE:	2000-09-08
71	PRIOR APPLICATION	NUMBER: 60/231,244
72	PRIOR FILING DATE:	2000-09-08
73	PRIOR APPLICATION	NUMBER: 60/233,066

```

; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,065
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,398

```

```

Query Match          26.1%; Score 1956; DB 5; Length 378;
Best Local Similarity 98.7%; Pred. No. 3,1e-85;
Matches 373; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY  997 ATYRTAIYONOHAMAKLELDPEMLYAHGPDLCRESDLRIAMANCEPALIGANYLEGS 1056
    1 ATYRTAIYONOHAMAKLELDPEMLYAHGPDLCRESDLRIAMANCEPALIGANYLEGS 60
QY  1057 LEEAKOLFGRLEFNDPDLREVWLNYPHPLQOEPTNDROLIETSPVLOKTEFEFAIGV 1116
    61 LEEAKOLFGRLEFNDPDLREVWLNYPHPLQOEPTNDROLIETSPVLOKTEFEFAIGV 120
QY  1117 ITHRYALLARATTLRTVGFNNHLLTGHNOEMELGDSIMQVATEYELFHPDHHGHLLTL 1176
    121 ITHRYALLARATTLRTVGFNNHLLTGHNOEMELGDSIMQVATEYELFHPDHHGHLLTL 180
QY  1177 LRSLSVNNRTOAKVAEELGMOEYAITNDKTRPVGLRRTKTLADLESFIALYTDKDEY 1236
    181 LRSLSVNNRTOAKVAEELGMOEYAITNDKTRPVGLRRTKTLADLESFIALYTDKDEY 240
QY  1237 VHTFMANVCFPRPKKEFIINQMDNDEKSQLQOCCFLTRTEGKEPDIPLYKTIQTVGSPSHAR 1296
    241 VHTFMANVCFPRPKKEFIINQMDNDEKSQLQOCCFLTRTEGKEPDIPLYKTIQTVGSPSHAR 300
QY  1297 YTVAVYFGERIGCGKSPSIQOAEEMGAMDALEKYNPOMAHOKRFTGKRYROELKEMR 1356
    301 YTVAVYFGERIGCGKSPSIQOAEEMGAMDALEKYNPOMAHOKRFTGKRYROELKEMR 360
QY  1357 WEREHOEREPDETEDIK 1374
    361 WEREHOEREPDETEDIK 378
DB

```

```

RESULT 6
US-10-103-313-307
; Sequence 307, Application US/10103313
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ207C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 307
; LENGTH: 263
; TYPE: PRT

```

```

; ORGANISM: Homo sapiens
US-10-103-313-307

```

```

Query Match          17.6%; Score 1320; DB 6; Length 263;
Best Local Similarity 97.7%; Pred. No. 2.9e-35;
Matches 251; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY  1118 FTHRYALLARATTLRTVGFNNHLLTGHNOEMELGDSIMQVATEYELFHPDHHGHLLTL 1177
    7 FTHRYALLARATTLRTVGFNNHLLTGHNOEMELGDSIMQVATEYELFHPDHHGHLLTL 66
DB  7 FTHRYALLARATTLRTVGFNNHLLTGHNOEMELGDSIMQVATEYELFHPDHHGHLLTL 66
QY  1178 RSLSVNNRTOAKVAEELGMOEYAITNDKTRPVGLRRTKTLADLESFIALYTDKDEY 1237
    67 RSLSVNNRTOAKVAEELGMOEYAITNDKTRPVGLRRTKTLADLESFIALYTDKDEY 126
QY  1238 HTEFMANVCFPRPKKEFIINQMDNDEKSQLQOCCFLTRTEGKEPDIPLYKTIQTVGSPSHAR 1297
    127 HTEFMANVCFPRPKKEFIINQMDNDEKSQLQOCCFLTRTEGKEPDIPLYKTIQTVGSPSHAR 186
QY  1298 YTVAVYFGERIGCGKSPSIQOAEEMGAMDALEKYNPOMAHOKRFTGKRYROELKEMR 1357
    187 YTVAVYFGERIGCGKSPSIQOAEEMGAMDALEKYNPOMAHOKRFTGKRYROELKEMR 246
QY  1358 WEREHOEREPDETEDIK 1374
    247 WEREHOEREPDETEDIK 263
DB

```

```

RESULT 7
US-09-547-599C-3281
; Sequence 3281, Application US/09547599C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lactrix, Bruno
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins
; FILE REFERENCE: GEN-1119C1
; CURRENT APPLICATION NUMBER: US/09/547,599C
; CURRENT FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 08/905,223
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,135
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,051
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,144
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,279
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/904,468
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,134
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,133
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 3475
; SOFTWARE: PatentL.pm
; SEQ ID NO 3281
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo Sapiens
; OTHER INFORMATION: kidney
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -23, -1
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.90
; FEATURE:
; OTHER INFORMATION: seq TIKFLILQKSNA/KR
; NAME/KEY: UNSURE
; LOCATION: 3

```

OTHER INFORMATION: Xaa = Ala,Pro  
US-09-547-599C-3281

Query Match 6.6%; Score 496; DB 5; Length 115;  
Best Local Similarity 95.7%; Pred. No. 1.3e-16;  
Matches 88; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 511 KKAHPRLHDELWYNDPGQNDGPKCSAKARTGIRHSIYFGEBAIRKPRPTNNAGR 570  
DB 24 KXNRLRLHDELWYNDPGQNDGPKCSAKARTGIRHSIYFGEBAIRKPRPTNNAGR 83  
QY 571 LEHRTTVSPPTNFTLDRPVTIYDDHEIYFE 602  
DB 84 LEHRTTVSPPTNFTLDRPVTIYDDHEIYFE 115

RESULT 8  
US-09-791-537-130915

Sequence 130915, Application US/09791537

GENERAL INFORMATION:

APPLICANT: Blonomix, Inc.

APPLICANT: Debe, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09/791,537

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 153055

SOFTWARE: PatentIn version 3.0

SEQ ID NO 130915

LENGTH: 249

TYPE: PRT

ORGANISM: Bacillus subtilis

US-09-791-537-130915

Query Match 4.2%; Score 315; DB 5; Length 249;  
Best Local Similarity 34.7%; Pred. No. 1.2e-07;  
Matches 85; Conservative 46; Mismatches 96; Indels 18; Gaps 7;

QY 1104 LOKLTFEEBAIGVITTHVRLARAFTLRTVGFNHLTLGH--NORMEFLGDSIMOLVATEY 1161  
DB 17 VEOFKFEQERISVHQNKEKLLYQAFTHSSYVNEHKKRPYEDNERLEFLGDVLELTISR 76  
QY 1162 LEFHPPDHHEGLTLRSSLVNNRTQAKVAEELGMOEYAITNDKTKRPVG-LRTKTLADL 1220  
DB 77 LFPKYPAMSEGDITKLRAAIVCEPSLVLAHLSFGDLVLGKGBEMTGKRRPALLADV 136  
QY 1221 LESFIALYTDKDELYVTFMNVCFPRLKFEFLINQDMDPKSLOQCCLTLRTEGK--- 1277  
DB 137 FFAFIALYLDQGLPEVSPFLKVVFPKINDGAPRHV-DRKSQLOE---YVQDQKGS 192  
QY 1278 EPDILYKTLQTVGSPSHARTYVAAYFKGERIGCGKPSIOAEMGAMDALEKYNF 1337  
DB 193 E-----YKISNEKGPANHREFEALVSLKGPELGVNGRSKKEAEOHAOEAALAKL---OK 244  
QY 1338 AHOKR 1342  
DB 245 HHTKQ 249

RESULT 9

US-09-791-537-62108

Sequence 62108, Application US/09791537

GENERAL INFORMATION:

APPLICANT: Blonomix, Inc.

APPLICANT: Debe, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09/791,537

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 62108  
LENGTH: 266  
TYPE: PRT  
ORGANISM: Bacillus subtilis  
US-09-791-537-62108

Query Match 4.1%; Score 309; DB 5; Length 266;  
Best Local Similarity 34.6%; Pred. No. 2.5e-07;  
Matches 84; Conservative 47; Mismatches 94; Indels 18; Gaps 7;

QY 1104 LOKLTFEEBAIGVITTHVRLARAFTLRTVGFNHLTLGH--NORMEFLGDSIMOLVATEY 1161  
DB 17 VEOFKFEQERISVHQNKEKLLYQAFTHSSYVNEHKKRPYEDNERLEFLGDVLELTISR 76  
QY 1162 LEFHPPDHHEGLTLRSSLVNNRTQAKVAEELGMOEYAITNDKTKRPVG-LRTKTLADL 1220  
DB 77 LFPKYPAMSEGDITKLRAAIVCEPSLVLAHLSFGDLVLGKGBEMTGKRRPALLADV 136  
QY 1221 LESFIALYTDKDELYVTFMNVCFPRLKFEFLINQDMDPKSLOQCCLTLRTEGK--- 1277  
DB 137 FFAFIALYLDQGLPEVSPFLKVVFPKINDGAPRHV-DRKSQLOE---YVQDQKGS 192  
QY 1278 EPDILYKTLQTVGSPSHARTYVAAYFKGERIGCGKPSIOAEMGAMDA---LEKYNF 1334  
DB 193 E-----YKISNEKGPANHREFEALVSLKGPELGVNGRSKKEAEOHAOEAALAKLKHHT 247  
QY 1335 POM 1337  
DB 248 KQL 250

RESULT 10

US-09-791-537-104336

Sequence 104336, Application US/09791537

GENERAL INFORMATION:

APPLICANT: Blonomix, Inc.

APPLICANT: Debe, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09/791,537

CURRENT FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 153055

SOFTWARE: PatentIn version 3.0

SEQ ID NO 104336

LENGTH: 266

TYPE: PRT

ORGANISM: Bacillus subtilis

US-09-791-537-104336

Query Match 4.1%; Score 305; DB 5; Length 266;  
Best Local Similarity 34.2%; Pred. No. 3.8e-07;  
Matches 83; Conservative 48; Mismatches 94; Indels 18; Gaps 7;

QY 1104 LOKLTFEEBAIGVITTHVRLARAFTLRTVGFNHLTLGH--NORMEFLGDSIMOLVATEY 1161  
DB 17 VEOFKFEQERISVHQNKEKLLYQAFTHSSYVNEHKKRPYEDNERLEFLGDVLELTISR 76  
QY 1162 LEFHPPDHHEGLTLRSSLVNNRTQAKVAEELGMOEYAITNDKTKRPVG-LRTKTLADL 1220  
DB 77 LFPKYPAMSEGDITKLRAAIVCEPSLVLAHLSFGDLVLGKGBEMTGKRRPALLADV 136  
QY 1221 LESFIALYTDKDELYVTFMNVCFPRLKFEFLINQDMDPKSLOQCCLTLRTEGK--- 1277  
DB 137 FFAFIALYLDQGLPEVSPFLKVVFPKINDGAPRHV-DRKSQLOE---YVQDQKGS 192  
QY 1278 EPDILYKTLQTVGSPSHARTYVAAYFKGERIGCGKPSIOAEMGAMDA---LEKYNF 1334  
DB 193 E-----YKISNEKGPANHREFEALVSLKGPELGVNGRSKKEAEOHAOEAALAKLKHHT 247



Db 643 SDKLSKRSVHRHGGQMSVENSEGRSPVSKVKD-SEQVEKENSDLDANLSCDSKD 701  
QY 455 ---ROEKAKAARPPWEPKTKLDELDLESSSECEDESDTSCSSSEVFDVIAETIKRK 511  
Db 702 TIRHQIKDKNRR-----KNK-----RSSREVSDD-----NGSSDSV-DDRKEAKRR 744  
QY 512 KAHPRDLDELWYNDPGQNDGFLCKCSAKARRGTGRH 549  
Db 745 RKEEKKTRKE-----EKKRRREERH 764

RESULT 13  
US-09-935-625-17076  
; Sequence 17076, Application US/09935625  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE  
; FILE REFERENCE: 2750-1481P  
; CURRENT APPLICATION NUMBER: US/09/935,625  
; CURRENT FILING DATE: 2001-08-24  
; NUMBER OF SEQ ID NOS: 33136  
; SEQ ID NO 17076  
; LENGTH: 847  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: 1..847  
; OTHER INFORMATION: Ceres Seq. ID no. 2708207  
US-09-935-625-17076

Query Match 3.6%; Score 273; DB 5; Length 847;  
Best Local Similarity 24.8%; Pred. No. 4.6e-05;  
Matches 158; Conservative 54; Mismatches 216; Indels 210; Gaps 32;

QY 8 HRMSFHRGCGCPRGCGHGAPSPAFSR-----POMLRLHPQP-----PVQYQYE 54  
Db 261 HRRPTHEGRROSPAPSRRRRSPSPARRRRSPSPARRRRSPSPARRRRSPTPPARQRR 320  
QY 55 PRSAPSTFNSNPAFLPRPDVFPFP-----PMPSAQGLPCCPI-----RPPFPN 104  
Db 321 SPSPARRHRSP-----PARRRRSPSPARRRRSPSP-ARRRRSPSLYRRNRSPSL 374  
QY 105 HQM-----RHPEVPVPCFPMPMPMPCPN-----NPPV-----PGAPR 137  
Db 375 YRNRGRSPPLAKRGSDSPVARRLDPTGARRLPSPSTEQRLPSPVQAQLPSP 434  
QY 138 GCGTFPFMMPPSPMHPPPP-----VMPQVNTQYPPGYSH--HNPFPFSFNQNNPS 190  
Db 435 RRAGLP--SPPAQRLPSPPRRAGLPSPMRI-----GSHAAHNLSPS-----PS 479  
QY 191 SFLPSANNSSPHFRLPYPLPKAPSERKSPERLKHVDHHRHDSH-----G 239  
Db 480 SLSPGGRK-----KVLSPVPRRRRSLTPDEERVSLSGGRTSPSHIKODGSMSPYRG 533  
QY 240 RGERHRSLLDRERGRSPDRRQDSRYRSDYDGRTPSRHRSYERSERERERHRRHNDNR 299  
Db 534 RGSKSPS-SRHQAKASPVARRSP-----TPVNRRS-RRSSASARSDDRR--RR 578  
QY 300 SPFLERSYKKEKKRSGRSTGLSVPEPACITPELGEIITKNTDSMAPPLEIVNHRSPS-R 358  
Db 579 SPSSSSPSR-----SRSP--VLHRSPPSR 602  
QY 359 EKKRARWEEDKRWSDNO---SSGDKNTSTIKEKEPETMDK-----NEEEEEE 406  
Db 603 GRKHORERRSPGRSLSEODRVONSKLLKRTSVPTDKRKQLEKLEVGREYHKEQERK 662  
QY 407 LKRPWIRCTHSENYSSDPMDQVGDSTVVGTSRLRLDYDKFEELGS-----454  
Db 663 SDKLSKRSVHRHGGQMSVENSEGRSPVSKVKD-SEQVEKENSDLDANLSCDSKD 721

QY 455 ---ROEKAKAARPPWEPKTKLDELDLESSSECEDESDTSCSSSEVFDVIAETIKRK 511  
Db 722 TIRHQIKDKNRR-----KNK-----RSSREVSDD-----NGSSDSV-DDRKEAKRR 764  
QY 512 KAHPRDLDELWYNDPGQNDGFLCKCSAKARRGTGRH 549  
Db 765 RKEEKKTRKE-----EKKRRREERH 784

RESULT 14  
US-09-935-625-17480  
; Sequence 17480, Application US/09935625  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPA  
; FILE REFERENCE: 2750-1481P  
; CURRENT APPLICATION NUMBER: US/09/935,625  
; CURRENT FILING DATE: 2001-08-24  
; NUMBER OF SEQ ID NOS: 33136  
; SEQ ID NO 17480  
; LENGTH: 859  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: 1..859  
; OTHER INFORMATION: Ceres Seq. ID no. 3023746  
US-09-935-625-17480

Query Match 3.6%; Score 273; DB 5; Length 859;  
Best Local Similarity 24.8%; Pred. No. 4.7e-05;  
Matches 158; Conservative 54; Mismatches 216; Indels 210; Gaps 32;

QY 8 HRMSFHRGCGCPRGCGHGAPSPAFSR-----POMLRLHPQP-----PVQYQYE 54  
Db 273 HRRPTHEGRROSPAPSRRRRSPSPARRRRSPSPARRRRSPSPARRRRSPTPPARQRR 332  
QY 55 PRSAPSTFNSNPAFLPRPDVFPFP-----PMPSAQGLPCCPI-----RPPFPN 104  
Db 333 SPSPARRHRSP-----PARRRRSPSPARRRRSPSP-ARRRRSPSLYRRNRSPSL 386  
QY 105 HQM-----RHPEVPVPCFPMPMPMPCPN-----NPPV-----PGAPR 137  
Db 387 YRNRGRSPPLAKRGSDSPVARRLDPTGARRLPSPSTEQRLPSPVQAQLPSP 446  
QY 138 GCGTFPFMMPPSPMHPPPP-----VMPQVNTQYPPGYSH--HNPFPFSFNQNNPS 190  
Db 447 RRAGLP--SPPAQRLPSPPRRAGLPSPMRI-----GSHAAHNLSPS-----PS 491  
QY 191 SFLPSANNSSPHFRLPYPLPKAPSERKSPERLKHVDHHRHDSH-----G 239  
Db 492 SLSPGGRK-----KVLSPVPRRRRSLTPDEERVSLSGGRTSPSHIKODGSMSPYRG 545  
QY 240 RGERHRSLLDRERGRSPDRRQDSRYRSDYDGRTPSRHRSYERSERERERHRRHNDNR 299  
Db 546 RGSKSPS-SRHQAKASPVARRSP-----TPVNRRS-RRSSASARSDDRR--RR 590  
QY 300 SPFLERSYKKEKKRSGRSTGLSVPEPACITPELGEIITKNTDSMAPPLEIVNHRSPS-R 358  
Db 591 SPSSSSPSR-----SRSP--VLHRSPPSR 614  
QY 359 EKKRARWEEDKRWSDNO---SSGDKNTSTIKEKEPETMDK-----NEEEEEE 406  
Db 615 GRKHORERRSPGRSLSEODRVONSKLLKRTSVPTDKRKQLEKLEVGREYHKEQERK 674  
QY 407 LKRPWIRCTHSENYSSDPMDQVGDSTVVGTSRLRLDYDKFEELGS-----454  
Db 675 SDKLSKRSVHRHGGQMSVENSEGRSPVSKVKD-SEQVEKENSDLDANLSCDSKD 733  
QY 455 ---ROEKAKAARPPWEPKTKLDELDLESSSECEDESDTSCSSSEVFDVIAETIKRK 511  
Db 734 TIRHQIKDKNRR-----KNK-----RSSREVSDD-----NGSSDSV-DDRKEAKRR 776

```

0Y 512 KAHPRDLHDELMYNDEGOMNDGPICKCSAKARRTGIRH 549
      : | | |
Db 777 RKEEKKTRK-----EKRRRREERH 796

```

```
Db 797 RKEKKTKRKE-----EKKRRREERH 816
Search completed: October 9, 2002, 18:35:10
Job time : 206 secs
```

RESULT 15  
US-09-935-625-17479

; Sequence 17479, Application US/09935625

; GENERAL INFORMATION

APPLICANT: N. ALEXANDROV et al.

TITLE OF INVENTION: POLYNUCLEOTIDES, CELLS, AND METHODS THEREOF CAPABLE OF MODULATING VARIOUS RESPONSES

```

; FILE OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P

```

CURRENT APPLICATION NUMBER: US/09/935,625

CURRENT FILING DATE: 2001-08-24

; NUMBER OF SEQ ID NOS: 33136

SEQ ID NO 17479

LENGTH: 879  
TYPE: PBT

TYPE: PRI

## FEATURES:

```

;
;  NAME/KEY: peptide
;

```

LOCATION

OTHER INFORMATION ;

US-09-935-625-17479

[illegible]

US-09-935-625-17479

Query Match	3.68	Score 273	DB 5	Length 879
-------------	------	-----------	------	------------

```

QY      8 HRMFPHGRCGRPRGGHGHGARSAPSR-----PQMLRLHPQD-----PVOGYXE 54
Db      293 HRPRTHGROSPARSPRRRRSSPPARRRRSSPPARRRRSSPPARRRRSSPPARRRRSSPPARRRR 352
QY      55 PPSAPSTTFNSPAPNPLPRPRDFVPPP-----PMPSAGPLPPCI---RDPFBN 104
Db      353 SPSPPARHRHSP-----PARRRSSPPARRRRSSPP-ARRRRSSPLYRRNRSPSL 406
QY      105 HON-----RHPPVPCPPMPMPCPN-----NPPV-----GAPP 137
Db      407 YRRNRSPPLAKGRKSRSPGSRSPVARLDDPGARLSPSITQRLSPSPVAORLSPPP 466
QY      138 GQGTFPEPMPPSPMHPPPP-----VMPQVNYQYPPGYSH--HNFPSPFSGPNSS 190
Db      467 RRAGLP--SPPPAORLSPSPPRAGLSPSPRI-----GSHANHLSPS-----PS 511
QY      191 SFLSANNSSPHRHLPYPLPKAEBERSPERLKHYYDHRRHSH-----G 239
Db      512 SLSPPGRK-----KVLSPVRRRRRLTDEEVSLSQGGRTTSPSHLKODGSMSPVNG 565
QY      240 RGERHRLDRERGRSPDRRRKODSRYSDDYDKRGTPSRHNSYRSRERERERHRHRDNRR 299
Db      566 RGSKSPS--SRHQKARSPVRRSP-----TPVNRSS--RRSSASRSPDRR--RR 610
QY      300 SPSLERSTKKKKRSGSYGLSVYPPBPACTPELPGLIKNTDSMAPPLEIYVHRSPS--R 358
Db      611 SPSSSRSPSR-----SRSP--VLHRSPSR 634
QY      359 EKKARMEEEKDRMSDNQ---SSGKDKNYTSIKEPEEETMPDK-----NEEEBE 406
Db      635 GRKHOREGRSGRLSEODRQVONSKLKRISYVDJTKRKRLPKLLEVGVEHYKQDEK 694
QY      407 LKFPWIRCTHSENYSSDPMDQVGDSTVYGSRLRDLYKFEELGSS----- 454
Db      695 SDKLSERKSVYRHGSOMSPEVENSEGRSPVSKYVD--SEOVYKEENSLDANLSCSDKD 753
QY      455 ---BOEKAKARPPWEPKTKLDEDLSESSCESOEDSTCGSSSDSEVFYATETKR 511
Db      754 TIRHOLKKNRR-----KNK-----RSSREVSDD-----NGSSSDV--DDRKAARR 796
QY      512 KAHPRDLHDELMYNDPGQMDGPLCKCSAKARRTGIRH 549

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 18:20:28 ; Search time 54 Seconds  
(without alignments)

2444.940 Million cell updates/sec

Title: US-09-900-425a-2

Perfect score: 7500

Sequence: 1 MMGNTCHRMSPHFGRCPR.....KRWEREHQREPDDEDIKK 1374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR-71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	890	11.9	412	2 T21419	hypothetical prote
2	558	7.4	682	2 T21420	hypothetical prote
3	316	4.2	551	2 S57447	HPBRII-7 protein
4	315	4.2	249	2 B69693	ribonuclease III (
5	307.5	4.1	229	2 AE1300	ribonuclease III h
6	300	4.0	229	2 AE1672	ribonuclease III h
7	293	3.9	1560	2 T42727	proliferation pote
8	285.5	3.8	225	2 B82073	ribonuclease III v
9	282	3.8	263	2 A83961	ribonuclease III r
10	281.5	3.8	232	2 F98012	ribonuclease III (
11	279.5	3.7	232	2 H95144	ribonuclease III (
12	276	3.7	948	2 A57640	retinoblastoma bin
13	275	3.7	760	2 T06291	extensin homolog
14	274.5	3.7	1531	2 T48946	hypothetical prote
15	273.5	3.6	1006	2 G86292	hypothetical prote
16	273	3.6	891	2 G84693	probable proline-r
17	271.5	3.6	620	2 S06733	hydroxyproline-ric
18	270	3.6	226	2 B84959	ribonuclease III (
19	268.5	3.6	231	2 B86725	ribonuclease III (
20	267	3.6	224	2 A81260	ribonuclease III (
21	267	3.6	243	2 D89896	RNase III (impor
22	267	3.6	998	2 T30930	hypothetical prote
23	265.5	3.5	1002	2 T43336	carboxypeptidase C
24	262	3.5	226	1 NRECC3	ribonuclease III (
25	262	3.5	226	2 A91058	RNase III (impor
26	262	3.5	226	2 F85902	RNase III, ds RNA
27	259	3.5	322	2 S25299	extensin precursor
28	258	3.4	231	2 S76204	hypothetical prote
29	257.5	3.4	272	2 T35656	ribonuclease III -

30	256	3.4	246	2 H70187	ribonuclease III (
31	255.5	3.4	1638	2 A42091	transcription acti
32	255	3.4	226	2 AC0829	ribonuclease III (
33	254	3.4	1611	2 T38236	hypothetical prote
34	253.5	3.4	368	2 C29356	hydroxyproline-ric
35	253.5	3.4	1048	2 T31425	C-terminal domain-
36	252	3.4	226	2 AF0331	ribonuclease III (
37	251.5	3.4	240	2 F64602	ribonuclease III -
38	251.5	3.4	1201	2 G64441	unknown protein (l
39	250.5	3.3	727	2 C84534	hypothetical prote
40	249	3.3	225	2 T09964	extensin Cys15 pre
41	249	3.3	907	2 E96636	hypothetical prote
42	249	3.3	1877	2 T21861	hypothetical prote
43	245	3.3	230	2 G97115	dsRNA-specific rib
44	244.5	3.3	489	2 T11622	extensin class 1 p
45	243.5	3.2	708	2 D96711	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T21419  
hypothetical protein F26E4.13 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T21419  
R.Lighting, J.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: 219419  
A:Accession: T21419  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-412 <Full>  
A:Cross-references: EMBL:Z81070; PIDN:CA803005.1; GSPDB:GM00019; CESP:F26E4.13  
A:Experimental source: clone F26E4  
C:Genetics:  
A:Gene: CESP:F26E4.13  
A:Map position: 1  
A:Introns: 359/2; 400/3

Query Match 11.9%; Score 890; DB 2; Length 412;

Best Local Similarity 43.1%; Pred. No. 3e-40;

Matches 181; Conservativity 85; Mismatches 136; Indels 18; Gaps 7;

QY	944	TLINISRLGDDPPTPSRINHNHERLEFLGDAYVEFLTSVHLYLPPSLDEGSLATYRTAI	1003
DB	2	SLFNIRKGTSGGP---ILHNERLEYGDVAVELVSHHLYEMLTHHFEGLATYRTAL	57
QY	1004	VONQHLAMLAKKLEIDPFMLVAHGPDLCRESDLRHAMNCFEALIGAVYLEGSLAEAKOL	1063
DB	58	VQNRNATLAKNCRIDEMQYSHGADLVNAEFKALNANFAVMAALYIDGGLAPCVI	117
QY	1064	FGRLTF-NDPDLREVLNPLRPLDQEPNDROLIETSPVLQKITEPEBAIGVFTYVR	1122
DB	118	FSKAMGHQPVLEKWDHINEHEIKREDPOGDGDSFIPTLSTFAHEERIGIOFNIR	177
QY	1123	LLARATLTGVEFNHITLGHNQMERTLGSIMQVATEVLFHFDHGHGHTLLRSSLY	1182
DB	178	LAKATRTNRINPNDLTGKNORLEWLGDSVLDLVSDLYRRFYHHEGMSLRTSLV	237
QY	1183	NNRQKVAEELGMOEYATVNTNKTSPV---GLRTKTLADLSEFIALLYTDKDLVYHT	1239
DB	238	SNQTAVVCDDLGFTFVY-----KAPYKTPBELKDKRADVLEATIGALVYDRGTEHCRA	292
QY	1240	FMNVCFPLKEFELINQDNNDRKSQLOQCCLTLR-TEGNEPDIPLYKTLQVYGPSHARTY	1298
DB	293	FIRIVCPRLKRIEESKENDAKSHLOQWCLAMRPSSSEPMPEYRVLYGEGPNNRIF	352
QY	1299	TVAVYFKGRIGCGKSPSIOQAEMGA---AMDALEKYNPPOM-AHQKRFIGKRYQOEKE	1354
DB	353	KIAVYKGRKLASAESNVHKAELRYAETALANLESMSFSKAKANNNSMRRRLDQTS	412





A.: Ehrlich, S.D.; Emmerson, P.T.; Entlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallet  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koehler, P.; Konungstein, G.; Kirogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogihara, A.; Oudea, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schreter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
akech, M.; Tamakoshi, A.; Tanaka, T.; Teipstra, P.; Tognoli, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A63580; MUID:98040403  
A:Accession: B69693  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-249 <KUN>  
A:Cross-references: GB:299112; GB:AL009126; NID:g2633902; PIDN:CAB13466.1; PID:g2633965  
A:Experimental source: strain 168  
R:Oguro, A.; Kakeshita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.  
Gene 172, 17-24, 1996  
A:Title: The effect of Srb, a homologue of the mammalian SRP receptor alpha-subunit, on  
A:Reference number: J04819; MUID:96257247  
A:Accession: J04821  
A:Molecule type: DNA  
A:Residues: 1-78, 'K', 80-137, 'K', 139-170, 'P', 172-242, 'E', 244-249, 'LNPPYDSGFGYVCLRI' <GCU  
A:Cross-references: DDBJ:D04116; NID:g1389548  
A:Note: The sequence in Genbank entry D04116, release 106.0, (PID:g1389549) has 138-Glu  
A:Genetics:  
A:Gene: rncs; srb  
C:Superfamily: ribonuclease III; double-stranded RNA-binding repeat homology  
C:Keywords: hydrolase  
F:172-245/Domain: double-stranded RNA-binding repeat homology <DSR>

Query Match 4.2%; Score 315; DB 2; Length 249;  
Best Local Similarity 34.7%; Pred. No. 6.8e-10;  
Matches 85; Conservative 46; Mismatches 96; Indels 18; Gaps 7;

QY 1104 LQKTEFEALGVFTHTVRLARFRLTYGPNHNL--TLGHNMREFLGDSIMQVATEYLFHF 1166  
DB 17 VEOKEKOEHSYVHFOEKLKLYQAFTHTSSVYNEHRRKRYEDNRELFGLGAVLELVSDYLF 76  
QY 1162 LFHPDHHEGHTLLRSSLVNNNTQAKVAEELGMOEYATINDTKRPVGLR--TLADLLESF 1224  
DB 77 LFAKTPAMSGDLTKRAALYCEPSLVFAELHSLFGLVLLGKEBMTGGRKRPALLADV 136  
QY 1221 LESFIALYDKDLEVHTFMNVCFFPRLEKFIILQDMPDKSQLQCCLTLETEGK-- 1277  
DB 137 FEATIGALYLDQGLEPESFLKYVFPKINDGAFSHVM--DFKSQLDE--YVORDGKSL 192  
QY 1278 EPDPLKTLQTVGSPHARTYVAVYFKGERIGCGKPSIQOAEKMAAMALEKYNPQM 1337  
DB 193 E----KISNEKPAHNRFEALVSLKGEPLGVNGRSGKAEQAHQAQBALAKL--QK 244  
QY 1338 AHOKR 1342  
DB 245 HHTKQ 249

RESULT 5  
AEI300  
ribonuclease III homolog rncs [imported] - *Listeria monocytogenes* (strain EGD-e)  
C:Species: *Listeria monocytogenes*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AEI300  
R:Glasier, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlan, K.D.; Fsihl, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefl, J.; Kuhn, M.; Kunst, F.; Kurpkat, G.; Madueno, E.; Maltournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A:Title: Comparative genomics of *Listeria species*.

A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AEI300  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-229 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC99883.1; PID:g16411259; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: rncs  
C:Superfamily: ribonuclease III; double-stranded RNA-binding repeat homology

Query Match 4.1%; Score 307.5; DB 2; Length 229;  
Best Local Similarity 33.0%; Pred. No. 1.5e-09;  
Matches 76; Conservative 51; Mismatches 86; Indels 17; Gaps 7;  
QY 1109 EFEBAIGVFTHTVRLARFRLTYGPNHNL--TLGHNMREFLGDSIMQVATEYLFHF 1166  
DB 6 ELQESVGFDFKVVLELQKQAFHTSSVYNEHRRNVDRNRELEFLGAVLELVSDYLF 65  
QY 1167 PDHHEGHTLLRSSLVNNNTQAKVAEELGMOEYATINDTKRPVGLR--TLADLLESF 1224  
DB 66 PDMAEESHMTKRAALYCEPSLVFAEAVHFSKY--VRLGKEKAGGRTRPALADVESF 124  
QY 1225 IALYTDKDLKLEYVHTFMNVCFFPRLEK--EFTLQDMPDKSQLQCCLTLETEGKEDIP 1282  
DB 125 ICALYLDNGIDKRVYFLERVFPKIDAGAYLQTVDY---KTQLDEIV-----QDRDVL 175  
QY 1283 L-YKTLQTVGSPHARTYVAVYFKGERIGCGKPSIQOAEKMAAMDALEK 1331  
DB 176 IEXDILGETGPAHNRKAFDAQVIVNGVILGKSGRTRKQAEQSAQFAINK 225

RESULT 6  
AEI672  
ribonuclease III homolog rncs [imported] - *Listeria innocua* (strain C1p11262)  
C:Species: *Listeria innocua*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AEI672  
R:Glasier, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlan, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefl, J.; Kuhn, M.; Kunst, F.; Kurpkat, G.; Madueno, E.; Maltournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of *Listeria species*.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AEI672  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-229 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC97149.1; PID:g16414420; GSPDB:GN00178  
A:Experimental source: strain C1p11262  
C:Genetics:  
A:Gene: rncs  
C:Superfamily: ribonuclease III; double-stranded RNA-binding repeat homology

Query Match 4.0%; Score 300; DB 2; Length 229;  
Best Local Similarity 31.8%; Pred. No. 3.8e-09;  
Matches 76; Conservative 53; Mismatches 88; Indels 22; Gaps 8;  
QY 1109 EFEBAIGVFTHTVRLARFRLTYGPNHNL--TLGHNMREFLGDSIMQVATEYLFHF 1166  
DB 6 ELQESVGFDFQVONVELLQKQAFHTSSVYNEHRRNVDRNRELEFLGAVLELVSDYLF 65  
QY 1167 PDHHEGHTLLRSSLVNNNTQAKVAEELGMOEYATINDTKRPVGLR--TLADLLESF 1224  
DB 66 PDMAEESHMTKRAALYCEPSLVFAEAVHFSKY--VRLGKEKAGGRTRPALADVESF 124  
QY 1225 IALYTDKDLKLEYVHTFMNVCFFPRLEK--EFTLQDMPDKSQLQCCLTLETEGKEDIP 1282  
DB 125 ICALYLDNGIDKRVYFLERVFPKIDAGAYLQTVDY---KTQLDEIV-----QDRDVL 175  
QY 1283 L-YKTLQTVGSPHARTYVAVYFKGERIGCGKPSIQOAEKMAAMDALEKYNPQMAHQ 1340

Db 176 IEDTILGETPAPHNNAFDAQVIVNGVILCKGSGRTKKAQEOSAOFAIN-----OLTHR 229

## RESULT 7

TA2727

Proliferation potential-related protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 02-Sep-2000  
C:Accession: TA2727

R:Witte, M.M.; Scott, R.E.

Submitted to the EMBL Data Library, November 1998

A:Reference number: Z22246

A:Accession: TA2727

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1560 <WIT>

A:Cross-references: EMBL:U83913; NID:g385884; PID:g385885; PIDN:AAC72432.1

A:Experimental source: strain Balb/C

A:Genetics:

A:Gene: P2P-R

A:Function:

A:Description: involved in hnRNP association and Rb1 binding

C:Superfamily: RING finger homology

F:57-107/Domain: RING finger homology <RRN>

Query Match 3.9%; Score 293; DB 2; Length 1560;

Best Local Similarity 23.8%; Pred. No. 1e-07;

Matches 151; Conservative 72; Mismatches 200; Indels 212; Gaps 34;

Db 25 HGARPSA--PSFRPONLRILHPQDPVOYQYEPSPA-----PSTFNSPAPNPLP 73

Db 296 NARPGGGRPGWENH-N-KIGYLVSPQOIRRGERSCYRINGRHNSRSQRTQGP-LP 353

Db 74 PRDPFPPRPSPASQGLPCRLPRPPNQHMPPRPCCPRMPRCMPNPPVP 133

Db 354 ARPVVPPVPP-----PLXPPP-----HTLPLP--PGVLPQPSQGP--P 391

Qy 134 GAPPGQGTFFPMPPSPMHP-----PPVPMQOVVY----- 166

Db 392 GQPPAG---YVPPPGFPAPANISTACFSHPVPTAHNTMPQTQALLSEEFYRQK 448

Qy 167 -----QYF---PGYSHHFPSPNSFQNNPSSFLPSANSSSHFHLPPYLPKAPS 217

Db 449 RUKESKPEYSGSSYSRSLYDTSKRSYSTRSYSRSFSRSHSRSPYP----- 502

Qy 218 ERRSPERLKHNDHRRHDSHGRGHRSLDRERGRSPDRRODSRYKSDYD--RGRTPS 276

Db 503 -RRGKGSKNY---KSRKSHG---YHRS-----RSRPPRTYRHSRKSFOAFRGOSP 550

Qy 277 RHRSTYRSHRERER-HRHR-----DNKRSPSLERSYKKEK 313

Db 551 K-RNVPRG-ETREYFNRYREVPYDIKAYYGRSVDRDPDEFKERYREWKREWEK 608

Qy 314 SGKSYGLSVPPACGTFLPGLIKNTDSNAP---FLELYNHSPPREKKRAMEEK 369

Db 609 YKGYAVGAQPPSA-----NREDSPEPLRLPLNIRN--SPTRGRREDIYAGQ 655

Qy 370 DWSN-----OSSGDKNYTSTIKEREPEETMPD-----KNE---EE 403

Db 656 SHRNANLGNVEKSLSTDSHNAKNPASKSEKSENVGGDKGNKHKHRRRRDEKCEE 715

Qy 404 EEELKLPWI---RCHSENY-----YSSDPMQVQVDSVYVGRSR 440

Db 716 SESFNPGLGFKRCRSGSIDETKTDLTFVFPREDATPVROEPMO--AESITFKSV 773

Qy 441 LRDLYDKKEEELGSRQEKAKA-----ARPPWPKRLDEDLSSSEF---CES 487

Db 774 DKDKRKDKPKVKSDTKRKSQSGATARKDNVLFKSKGPOEKVDGDKREKSPSEPLKA 833

Qy 488 DEDST-----CSSSDSEVFVDIAIKRKAH 514

Db 834 KEATKIDSVKRSSSQKDEKVTG-----PRKAH 863

## RESULT 8

B82073

Ribonuclease III VC2461 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: B82073

R:Haideberg, J.F.; Eison, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of Both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: B82073

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-225 <HET>

A:Cross-references: GB:AE004316; GB:AE003852; NID:g9657034; PIDN:AAF95603.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2461

A:Map position: 1

C:Superfamily: ribonuclease III; double-stranded RNA-binding repeat homology

Query Match 3.8%; Score 285.5; DB 2; Length 225;

Best Local Similarity 32.2%; Pred. No. 2.2e-08;

Matches 75; Conservative 42; Mismatches 101; Indels 15; Gaps 5;

Db 1101 SPVLQTLFEFEAGVIFTHVRLARAFILRVGFNHLTLGNHQMREPLGDSIMQVARE 1160

Db 2 TPWNKLT---SKIGYFKETELINTALTRHSANGK-----HNERLEFLGDSITLSFVIAD 53

Qy 1161 YLFHPEDHGHGHTLRSLSLVNRRQAKYAEELGQDEVAIFNDKRRPVYGTKE--LND 1219

Db 54 ELTKRPKNEGDSRRKATLVKGNLTAEIGREFDLDYIKLGPGLKSGGFRSDILAD 113

Qy 1220 LLESFIALYTDKDLVYHFMVNCFFPRLKEPILNDQWMDPKSQLOOCCLTLPTEGKEP 1279

Db 114 AVEAIIAIVLDSLETARSLVLEWYHGRLEIKPGASQDPRTRIQEF-----LGRKK 168

Qy 1280 DIPYKTLQVGSFHARTTVANYFKG-ERIGCGKPSIOAMGMAADALEK 1331

Db 169 PLPYVTYINIKGAHNOEFTVACEVAGMDPPVIGKTSRKKAQAAATLBD 221

## RESULT 9

A83961

Ribonuclease III rncs [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: A83961

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: A83961

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-263 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06208.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: rncs

C:Superfamily: ribonuclease III; double-stranded RNA-binding repeat homology

Query Match 3.8%; Score 282; DB 2; Length 263;

Best Local Similarity 30.9%; Pred. No. 4.1e-08;

Matches 77; Conservative 50; Mismatches 110; Indels 12; Gaps 5;

Qy 1084 HPLQLEPNDRDLIEPVLQKLTQFEAEIGVIFHVLRLARFTLRTVGNH--LTIG 1141

Db 15 HSERRRP---KRLTLFAKQOQMFDELRLTLNLFANKRLVLAQFVHSSVYVNEHRIOSCK 71

**Query Match** 3.8% Score 281.5; DB 2; Length 232;  
Best Local Similarity 32.9%; Pred. No. 3.8e-08;  
Matches 79; Conservative 47; Mismatches 91; Indels 23; Gaps 9;

OY 1099 ETSPILOKLTPEEEAIGVIFTHRLARAFLTRVGFMH--LTLGHNCRMEFIGDSIMQL 1156  
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 3 ELQTVLKNNFEI-----FADKKILETAFTHTSTANERHLIKTISHNERLEFLGDAYVL 56  
OY 1157 VATEYLFIHPDHNEGHLLTRSSLVNNRTOAKVAEELGMDEYATINDTKRPVGLRTK- 1215  
: : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 57 LTSELYLKYKPKRPGEDSLKRAMIVRESLAGFARDQDFQ-ITLKGGEKSGGRNRD 115  
OY 1216 -TLADLLESFYALTYTDMDLEYVTFMNVCFPRRK--ELLNDQWNPDKSLOCCCTL 1272  
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 116 TILGAFFEFAGALLDDVDNAVKKEFIYQMIPKVEAGEFEFITDY--KTHLDK--LL 169  
OY 1273 RTEGKEPDIPL-YKTIQVWGSPSHARTYVAAYFKERGIRCGKPSIOQAEMGADADELK 1331  
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 170 QVNG---DVATIFYVISERGAHNDVFDEVYLVESKISIGCGGRSKKLAEDEAAKNAYEK 226

**RESULT 11**  
**H95144**  
ribonuclease III [imported] - Streptococcus pneumoniae (strain TIGR4)  
C.Species: Streptococcus pneumoniae  
C.Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C.Accession: H95144  
R.Teteljin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Hickley, M.R.; Radune, D.; Holtzapfe,  
nson, T.; Rickert, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A.Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison  
Article: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.

OY 1142 HNQRNFEIGDSIMOLVATEYLEFIHPDHNEGHLLTRSSLVNNRTOAKVAEELGMDEYAI 1201  
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 72 DNERLEFLGDAYLAVALASQYLYKAEEQMSSEGDMKLRASICEPSIALAEELHRGELVL 131  
OY 1202 TNDKRRPVG-VRTKTLADLLESFYALTYTDKDLEYVTFMNVCFPRRKPEFILNDQWND 1260  
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 132 LGKGEEMTGRRRAPLADVESPFGALXYLDQGMNAVLFLEERTIYPYSISGASH-RMD 190  
OY 1261 PKSQLOCCCLTRTEGKEPDIPLYKTIQTGVGSPSHARTYVAAYFKERGIRCGKPSIOQA 1320  
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 191 FKSQLQEF-----IQDRNLGIHIHEIVQERGPANHREFEVSELNNETLIGVTGRSKREA 245  
OY 1321 EMGAMDAL 1329  
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 246 EDHAHQDAL 254

**RESULT 10**  
**F98012**  
ribonuclease III (EC 3.1.26.3) [imported] - Streptococcus pneumoniae (strain R6)  
C.Species: Streptococcus pneumoniae  
C.Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
R.Hockings, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E  
y, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A.Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A.Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A.Reference number: A97872; MUID:21429245; PMID:11544234  
A.Accession: F98012  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1232 <KUR>  
A.Cross-references: GB:AEO7317; PIDN:AAK99930.1; PID:G15458754; GSPDB:GN00174  
A.Genetics:  
A.Gene: rncS  
C.Superfamily: ribonuclease III, double-stranded RNA-binding repeat homology

[illegible]

```

Db 472 LKEEKKKKSKLDEFINDPAKELMEKKKIOKERRRBFSSKSPYSSSSYSRSSYTTXSKGRS 531
OY 185 F0NNDSFLPFSANNNSSPHFRHLLPPYLPKAPSERRRSPERLKHYYDHRHDSHGGERH 244
Db 532 GSTRSRSTRSRSSRSHSSSRSSYKSPYP-----RRGRKSSNY--RSRSRSHG--YH 578
OY 245 RSLDRREGSRSDRRRQGRSYKSDYD-RGRTPSRHSRHSRHSRERER----- 289
Db 579 RS-----RSRSPYRRYSRSRSPQAFRGQSSKNNKNNVDGETEREYFNRYREVPYDMK 633
OY 290 -----ERHRHRRNRRSPSLERSYKKEYKRGSRGSLGVBPACCTPELGE 336
Db 634 AYYGSHVDFRDFEKEKERE-----WERKYREWEXKYYKGYAQAQRPSSA----- 679
OY 337 IIKNTDSNAP-----PLEIVNHRSPSRKKRARWEBEKDRWSDNOSG-----KD 381
Db 680 ---NENENSPERFLPLNIRN--SPFTRGRRREDDYVGQGSHRSNISNYPPEKLASDGHNQ 734
OY 382 KNYTSIKKEPEETMPD-----KNBEEBELLKPWINGCTSENYSDDPM-- 427
Db 735 KDNITSKKKESENNAPGDKGNKKHKHRRKRRKGEESEGFNLPLETTSRK---SKEPGV 790
OY 428 -DOVGDSVTYVTSR-----LRDIYDKFEE--LGSROEKAKAARPPWPEPTKIDLE 478
Db 791 EENKTDLSFLVPSRODATPVRD--EPMDAESITFKSVSEKOKRER--DKPRAKDKTKR 845
OY 479 SSSBECESDESDTSCSSSSDSE-----VFDVIAETIKKKAHPDRL 518
Db 846 KNDGSAVSKENIYPAKGPQEKVDGVDYDLIDLMLQJLKKPERRRLRL 893

```

```

RESULT 13
T06291
    extensin homology T9E8.80 - Arabidopsis thaliana
C:/Species: Arabidopsis thaliana (mouse-ear cress)
C:/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-Jun-1999
A:/Accession: T06291
R:/Beyan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue
submitted to the Protein Sequence Database, April 1999
A:/Reference number: Z15588
A:/Accession: T06291
A:/Molecule type: DNA
A:/Residues: 1-760 <BEV>
A:/Cross-references: EMBL:AL049608
A:/Experimental source: cultivar Columbia; BAC clone T9E8
C:/Genetics:
A:/Map position: 4
A:/Note: T9E8.80

```

Query Match	3.78;	Score 275;	DB 2;	Length 760;
Best Local Similarity	35.18;	Pred. No. 3.8e-07;		
Matches 88;	Conservative 15;	Mismatches 96;	Indels 52;	Gaps 12;

[illegible]

RESULT 14  
 T48946  
 Hypothetical protein T15B3.60 - *Arabidopsis thaliana*  
 C|Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C|Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C|Accession: T48946  
 R|Jordan, N.; Banerly, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.  
 submitted to the Protein Sequence Database, April 2000  
 A|Reference number: 225009  
 A|Accession: T48946  
 A|Status: preliminary  
 A|Molecule type: DNA  
 A|Residues: 1-1531 <TOR>  
 A|Cross-references: EMBL:AL163975; GSPDB:GN00061; ATSP:T15B3.60  
 A|Experimental source: cultivar Columbia; BAC clone T15B3  
 A|Genetics:  
 C|Gene: ATSP:T15B3.60  
 A|Map position: 3  
 A|Intons: 45/2; 100/3; 138/3; 183/3; 227/3; 278/3; 322/1; 357/2; 421/3; 475/2; 584/1

Query Match	3.7%	Score 274.5	DB 2	Length 1531
Best Local Similarity	21.3%	Pred. No. 9.7e-07		
Matches 247	Conservative 142	Mismatches 420	Indels 351	Gaps 58

QY 372 WSDNQSSGKDKNYT\$IKKEPEETMPDKNEEEEEELKPVWIRCTHSENYSSDPMDQVG 431

Db 455 YSQQKHAKQNSKSI~~ML~~-~~ERG~~NP~~KQ~~RDHL~~HD~~LM~~RE~~VL~~IQ~~DP~~EA~~PN~~LK~~SC~~PP~~VKN~~G~~ 512

0V 432 DST-VGTSRLDLYDKFEELGSR0EKAKARPPWE---PP-KTKLDED-LESSSES 483

Dh 513 HGVKEIGSMVIPDSNITYVSEEA<sup>1</sup>STOTMSD<sup>2</sup>---PPSRNEOLPPCKKLR<sup>3</sup>LDNNLL<sup>4</sup>OSNGKE 569

00 484 ECESDEDSTCSSSSDSEVEDYIAETKKKAHPDRI.HDET.---WYNPDGOMNDGP----- 534

Db 570 KVA SSKSSSAAGSK-----KRFETHTTCANALSGTW-----GENTDGA TEOAYK 617

04 535 ---TCKCSBK-----ABPTG-TPHSTYBCEFAIK----BCBPM----- 564

618 EDECCNISCCEVSCSCTTSCSTAEVDCVCEMDYIVBRTVKASVSDCCOTPI\$CFFETVK 677

-----  
C-2  
  
E C E  
  
-----  
THUWANT TUITTWCDDDDWWER MDDDDWUTREDDDUUEVTEEECEGUAUADIEMWT 615

[illegible]

\_\_\_\_\_

[illegible]

—

[illegible][illegible][illegible]

DB 84 / --KKYGLVLAH--PNDPLMKLKQSHAHNLVDENEEMVKTJEPKAGNVKRRKKNL---- 898

QY 845 QGFWKIGLRSDVCQHAMLPRVLTNHLRYHQCLMHLBDKLLGYTFQDKRLQLDAMTHPS--H 90

Db 899 -----HAHLPEL-----LARID-----VRÄVLKSIYLLPSVMH 926

02 903 HLNFGMPDHARNSLNCGIRQPKYGDRKVHHMMRRKKGINTLINMSRLGQDDPTPSRI 962

```
Db 929 RLESIMLASQREI-DCSIDNFSISSTSI-----LEAVTTL-----TCPES 969
```

QY 963 NHNERLEGDAVEFLTSHLYLFPSEEGGLATYRTAIVONQHLAML--AKKLE--- 101

Db 970 F5MERLELLGDSVLKYVASCHLEFKYPDKDEGLSRQRQSIISNSNLHRLTTSRKLQGYI 102

QY	1018	----	LDPEMLAHG-----	PDLC-----	-----	RESDLR-----	1038
Db	1030	ENGAFEPBRMTAPGOSFLPYVPCCKGIDTREVPLDPRKFTFENMTIKICKSCDMGRWVVS					1089
QY	1039	-AMANCEFALGAVYLKESGLEBEAKOLFGRLLFN--	-DPDLREVWLVNLYLHPLDLOEPNTD				1094
Db	1090	KVSQDCAEALIGAVYVSGLSASLHMNMKLGIDVDFDNLVVEAINRY-	SIRCYTPRED				1147
QY	1095	ROLIETSVLOKLTPEFEALIGVITFHVALLARAFILKRTVGNHNLTLGN--	QOREFGLD				1151
Db	1148	-----	ELIELEKRIIDHEFSAKFLKEALT--	-----	HSLRESYSYERLEFGLD		1189
QY	1152	SIMOLVATEYELFIFHPDHEGHLTLRSSLVNNRTOAKVAE-----	ELGM				1196
Db	1190	SVLDPLIRHLFENLYEQGPREMTDLRSACVNNENFEQAVAKNNLHTLOKCATVLEQOI					1249
QY	1197	QEYALT--	NDKTKR--	PVGLRTKTLADLSEFIALATYDXOLEVHTFPMVCEFPRLKE			1251
Db	1250	NDYLMSEFOKPBETGRSIPSIQGRKALGDVSEISGALLIDRILDQYWR--	VEEPILSP				1307
QY	1252	FLINDWM--	DKPSOLOQCCLTL--	RTEGKEPDPIELKTKLO--	TYGSPHARTYT		1299
Db	1308	LVTPEKTLDPYRLRELNECLDSLGTFPRVKCSNDGYKKAOTIOLOLDVLLTGDGSEOT--					1365
QY	1300	VAVYFKGERIGCGKPSIQQAEMGAAMDALERYNEPMAHOKRFIGRY--	ROELKEM				1355
Db	1366	-----	NKALMGKAAASHLNLQ-----	-----	EKRNRKRTSLGDNQSSNDV		1401
QY	1356	RWEREHOERE--	PDLETEDIK				1373
Db	1402	NLACNHSRDREMLTSETTEIIO					1421

RESULT 15  
G86292  
hypothetical protein F7H2.17 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence-revision 02-Mar-2001 #text-change 31-Dec-2001  
C:Accession: G86292  
R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malli, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: G86292  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1006 <SNO>  
A:Cross-references: GB:AE005172; NID:g8927662; PIDD:AA82153.1; GSPDB:GN00141  
C:Genetics:  
C:Map position: 1

Query Match	3.68;	Score 273.5;	DB 2;	Length 1006;
Best Local Similarity	32.98;	Pred. No. 6.4e-07;		
Matches 99;	Conservative 17;	Mismatches 98;	Indels 87;	Gaps 19

[illegible][illegible]

Search completed: October 9, 2002, 18:26:49  
Job time : 58 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 17:49:53 ; Search time 88 Seconds  
(without alignments)  
2701.080 Million cell updates/sec

Title: US-09-900-425a-2

Perfect score: 7500  
Sequence: 1 MMGNCTCHRMSEFHRGCGCPR.....MRWERHOREPDEDIDIK 1374

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7486	99.8	1374	4	O9NR4 homo sapien
2	4071	54.3	769	4	O9NW73 homo sapien
3	2963	39.5	1327	5	O9XYN5 drosophila
4	2847	38.0	541	4	O9Y2V9 homo sapien
5	2811.5	37.5	1071	5	O960Y4 drosophila
6	1973	26.3	380	4	O9Y4Y0 homo sapien
7	1448	19.3	1086	5	O01327 caenorhabdi
8	1417	18.9	860	5	O9U908 caenorhabdi
9	711	9.5	134	11	O9CTG2 mus musculu
10	340	4.5	1151	5	O9YAY4 drosophila
11	316	4.2	551	4	O16630 homo sapien
12	306.5	4.1	588	4	O9BW18 homo sapien
13	300	4.0	229	16	O92AK3 listeria in
14	288.5	3.8	1591	11	P97868 mus musculu
15	285	3.8	409	10	O9SAM1 O9sbml volvox cart
16	285	3.8	478	4	O9BSJ7 homo sapien

17	285	3.8	652	5	O95TS9	drosophila
18	284.5	3.8	1616	4	O96PH3	homo sapien
19	282	3.8	263	16	O9KA05	bacillus ha
20	279.5	3.7	232	16	O97G66	streptococ
21	276	3.7	548	4	O15250	homo sapien
22	275.5	3.7	594	5	O9VEP4	drosophila
23	275	3.7	760	10	O9TOK5	arabidopsis
24	274.5	3.7	1531	10	O9LXW7	arabidopsis
25	273.5	3.6	1006	10	O9LMO1	arabidopsis
26	273	3.6	891	10	O9ZW08	arabidopsis
27	273	3.6	894	10	O9FTB2	arabidopsis
28	272.5	3.6	702	5	O9VSH4	drosophila
29	271.5	3.6	230	16	O9A105	strepococ
30	271	3.6	2994	5	O95ZG5	dictyostell
31	269	3.6	884	4	O00302	homo sapien
32	268	3.6	243	16	O9J1T1	staphylococ
33	267	3.6	224	16	O9PW40	campylobact
34	267	3.6	243	16	O99UN7	staphylococ
35	267	3.6	988	5	O01864	caenorhabdi
36	266	3.5	687	10	O948Y7	volvox cart
37	265.5	3.5	1002	3	O13849	schizosacch
38	265	3.5	740	3	O96U76	neurospora
39	261	3.5	763	2	O9XDH2	mycobacteri
40	261	3.5	1236	4	O9C012	homo sapien
41	259.5	3.5	555	10	O9FP06	chlamydomon
42	259	3.5	322	10	O09084	lycopersico
43	259	3.5	1217	4	O9ULI5	homo sapien
44	258	3.4	1400	11	O9ESU6	mus musculu
45	256	3.4	956	10	O9LJ64	arabidopsis

## ALIGNMENTS

RESULT 1					
ID	O9NR4	PRELIMINARY;	PRT;	1374 AA.	
AC	O9NR4;				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	RIBONUCLEASE III.				
GN	RN3.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20538440; PubMed=10948199;				
RA	Wu H., Xu H., Miraglia L.J., Crooke S.T.;				
RT	"Human RNase III is a 160-kDa protein involved in preribosomal RNA				
RT	processing."				
RL	J. Biol. Chem. 275:36957-36965(2000).				
DR	EMBL: AF189011; AAF80558.1; -.				
DR	InterPro: IPR001159; DS_RBD.				
DR	InterPro: IPR002865; P_Rich_extensn.				
DR	InterPro: IPR000999; RNase_3.				
DR	Pfam: PF00035; dsrm; 1.				
DR	Pfam: PF00636; Ribonuclease.3; 2.				
DR	PRINTS: PR01217; PRICHEXTNSN.				
DR	SMART: SM00358; DSRM; 1.				
DR	SMART: SM00355; RIBOC; 2.				
DR	PROSITE: PS50137; DS_RBD; 1.				
DR	PROSITE: PS00517; RNASE_3_1; 2.				
DR	PROSITE: PS50142; RNASE_3_2; 2.				
SO	SEQUENCE 1374 AA; 159244 MW; 888108523912F705 CRC64;				
Query Match					
Best Local Similarity 99.8%; Score 7486; DB 4; Length 1374;					
Matches 1372; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
Oy	1	MMGNCTCHRMSEFHRGCGC	GAARSA	SPRPNLRLTLHQDPVQYQYEP	PSAPS 60

```

Db 1 M0AGTCTCARMSPHGRGCGKRGSGHGAAPSASFRRQNLRLH0QDPVYQYUYPAPS 60
QY 61 TTFSNSAPNPLPRPDVPPPPMPPSAOGLPPCPLRPFPNOKRHPPVPPCPRM 120
Db 61 TTFSNSAPNPLPRPDVPPPPMPPSAOGLPPCPLRPFPNOKRHPPVPPCPRM 120
QY 121 PPMPCPNPPVPGAAPGQGTFFPMMPSPMHPPPPPVQOVYQYPRGYSHNPPPP 180
Db 121 PPMPCPNPPVPGAAPGQGTFFPMMPSPMHPPPPPVQOVYQYPRGYSHNPPPP 180
QY 181 SPNSONNPPSFLPSANSSSPHFRHLPPYPLKAPSRBRDELKHYDDHNRHDSHGR 240
Db 181 SPNSONNPPSFLPSANSSSPHFRHLPPYPLKAPSRBRDELKHYDDHNRHDSHGR 240
QY 241 GERHSLDRRERGSPDRRSDSYRSDYDRGRTPSRRHSYERERERERHRRDRNRS 300
Db 241 GERHSLDRRERGSPDRRSDSYRSDYDRGRTPSRRHSYERERERERHRRDRNRS 300
QY 301 PSLEHSYKKEYKRGSGSYGLSVBPAGCTPELPGELIKNTDSNAPPLEIVNHRSPSREK 360
Db 301 PSLEHSYKKEYKRGSGSYGLSVBPAGCTPELPGELIKNTDSNAPPLEIVNHRSPSREK 360
QY 361 KRAREEKKDMSDNOSGKCKNTSTIKERPEETMDPKNEEDELKPYWICTHSEN 420
Db 361 KRAREEKKDMSDNOSGKCKNTSTIKERPEETMDPKNEEDELKPYWICTHSEN 420
QY 421 YSSDPMDQVDSYVGTSLRLDLYDKFEELSGROEKAKAPRPMPPKTLDEDESS 480
Db 421 YSSDPMDQVDSYVGTSLRLDLYDKFEELSGROEKAKAPRPMPPKTLDEDESS 480
QY 481 SESECEDEDSTCSSSDSEVFDVIAETIKRKAHPDRHLDLWYNDPQMDGFLCKCSA 540
Db 481 SESECEDEDSTCSSSDSEVFDVIAETIKRKAHPDRHLDLWYNDPQMDGFLCKCSA 540
QY 541 KARRGTHSYTPEGEALIKPCRPMTNNGRLPHRTYSPPTNLTLRPVTEYDHEXI 600
Db 541 KARRGTHSYTPEGEALIKPCRPMTNNGRLPHRTYSPPTNLTLRPVTEYDHEXI 600
QY 601 FEGFSMFAHAPLTINPLCKVIRFNIDYTHFIEEMMPNFCVKLEFLSLFDFDILEY 660
Db 601 FEGFSMFAHAPLTINPLCKVIRFNIDYTHFIEEMMPNFCVKLEFLSLFDFDILEY 660
QY 661 DWNLTGKPLFEDSPCCPRFHPMPRVRLPDGKKEVLSMHQILLYLRCSKALVPEEIEA 720
Db 661 DWNLTGKPLFEDSPCCPRFHPMPRVRLPDGKKEVLSMHQILLYLRCSKALVPEEIEA 720
QY 721 NMLQWEELEMGKYAECKGMIVTNPGRKPSVRIIDQDREOFNDVITFPITVHGRIPA 780
Db 721 NMLQWEELEMGKYAECKGMIVTNPGRKPSVRIIDQDREOFNDVITFPITVHGRIPA 780
QY 781 QLSYAGDPOYOGLMKSYVKLHLLANSPPKVQTDOKLQAEELALOKIROKNTMRRETV 840
Db 781 QLSYAGDPOYOGLMKSYVKLHLLANSPPKVQTDOKLQAEELALOKIROKNTMRRETV 840
QY 841 ELSSGGEFKTGTGRSDYCOHAMLPVLTNHHIRYHOCIMHLDLIGTTFDRCCLDLAMTHP 900
Db 841 ELSSGGEFKTGTGRSDYCOHAMLPVLTNHHIRYHOCIMHLDLIGTTFDRCCLDLAMTHP 900
QY 901 SHHLLFGNMPDHARNSLSCNGIRQPKYGRKVVHNMHMKKGINTLINIMSLGODDDPPS 960
Db 901 SHHLLFGNMPDHARNSLSCNGIRQPKYGRKVVHNMHMKKGINTLINIMSLGODDDPPS 960
QY 961 RINHERLEFLGDAVEELTGYHLTYLPPSLEEGGLATYRAIYONCHILAMAKLELDP 1020
Db 961 RINHERLEFLGDAVEELTGYHLTYLPPSLEEGGLATYRAIYONCHILAMAKLELDP 1020
QY 1021 FMYLVAHGDLCRESDLRHAMANCEFALIGAVYLRGSSLEAKOLFGRLLFNDPDLREVWLN 1080
Db 1021 FMYLVAHGDLCRESDLRHAMANCEFALIGAVYLRGSSLEAKOLFGRLLFNDPDLREVWLN 1080
QY 1081 YPLHPLQLOEPNTDROLLETSPVLOKLEFEELAGVITTHRLAKARATLETGVPNHLTL 1140
Db 1081 YPLHPLQLOEPNTDROLLETSPVLOKLEFEELAGVITTHRLAKARATLETGVPNHLTL 1140

Db 1081 YPLHPLQLOEPNTDROLLETSPVLOKLEFEELAGVITTHRLAKARATLETGVPNHLTL 1140
QY 1141 GHNORMEFLGDSIMOLVATEYLFHPDPHHEGHLTLRSLVNNKPTQAKVAEELGMOEYA 1200
Db 1141 GHNORMEFLGDSIMOLVATEYLFHPDPHHEGHLTLRSLVNNKPTQAKVAEELGMOEYA 1200
QY 1201 IINDKTRPVGLKTKTLADLESFIALYTDKDELEVHFMVNCFFPRKERTLNQMDND 1260
Db 1201 IINDKTRPVGLKTKTLADLESFIALYTDKDELEVHFMVNCFFPRKERTLNQMDND 1260
QY 1261 PKSLOOCCLTLTREKEEDIPLYKTQTVGSPHARTYVAVYFGERIGCKGSPSIOQA 1320
Db 1261 PKSLOOCCLTLTREKEEDIPLYKTQTVGSPHARTYVAVYFGERIGCKGSPSIOQA 1320
QY 1321 EMGAAMDALERYNFPQMAHOKFETGRVYOBELKEMWEPHOREPDEDIDIK 1374
Db 1321 EMGAAMDALERYNFPQMAHOKFETGRVYOBELKEMWEPHOREPDEDIDIK 1374

RESULT 2
Q9NM73 PRELIMINARY; PRT; 769 AA.
ID 09NM73:
AC 09NM73:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOHERICAL 89.7 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRIO;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiraori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takasashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Nishimura K., Iwayanagi T.;
RT "NEO human cDNA sequencing project."
RL Submitted (Feb-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK001121; BAA91511.1; -.
DR InterPro: IPR001159; DS_RBD.
DR InterPro: IPR000999; RNase_3.
DR Pfam: PF00035; dsrm; 1.
DR Pfam: PF00636; Ribonuclease_3; 2.
DR SMART: SM00358; DSRM; 1.
DR SMART: SM00355; RTBC; 2.
DR PROSITE: PS00137; DS_RBD; 1.
DR PROSITE: PS00517; RNase_3_1; 2.
DR PROSITE: PS0142; RNase_3_2; 2.
SQ SEQUENCE 769 AA; 89706 MW; C92261C723C047FA CRC64;

Query Match 54.3%; Score 4071; DB 4; Length 769;
Best Local Similarity 99.5%; Pred. No. 3; Se-281;
Matches 765; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 606 MFAHAPLTINPLCKVIRFNIDYTHFIEEMMPNFCVKLEFLSLFDFDILEYDMLNK 665
Db 1 MFAHAPLTINPLCKVIRFNIDYTHFIEEMMPNFCVKLEFLSLFDFDILEYDMLNK 660
QY 666 GPLFEDSPCCPRFHPMPRVRLPDGKKEVLSMHQILLYLRCSKALVPEEIEANMLQW 725
Db 61 GPLFEDSPCCPRFHPMPRVRLPDGKKEVLSMHQILLYLRCSKALVPEEIEANMLQW 120
QY 726 EELEMGKYAECKGMIVTNPGRKPSVRIIDQDREOFNDVITFPITVHGRIPAQLSTA 785
Db 121 EELEMGKYAECKGMIVTNPGRKPSVRIIDQDREOFNDVITFPITVHGRIPAQLSTA 180
QY 786 GDPYOGLMKSYVKLHLLANSPPKVQTDOKLQAEELALOKIROKNTMRRETVELSSQ 845

```



Dd	181	GDPOYLAKMYSYKLRHLNLANSKVQGTQKOLQAGNEELQKILROKNNRREPVYELSSQ	240
Qy	846	GEWKTGIRSDVCQHAMLPLYLTHHIRYHOCMLHDLKLGITFODRCCLQLQAMTHPSHLN	905
Dd	241	GFMTGTIRSDVDCQHAMLPLYLTHHIRYHOCMLHDLKLGITFODRCCLQLQAMTHPSHLN	300
Qy	906	FGMNPDIARNSLNSCGIRQKYGDKRVYHHMHKRRKCIINFLINMSRLGDDPTSPRIMN	965
Dd	301	FGMNPDIARNSLNSCGIRQKYGDKRVYHHMHKRRKCIINFLINMSRLGDDPTSPRIMN	360
Qy	966	ERLEFLIDAVAEFLTSVHLKYLFPSPSLKEGGIATYPRATVQNOHMLAKKLEDFEMLYA	1025
Dd	361	ERLEFLIDAVAEFLTSVHLKYLFPSPSLKEGGIATYPRATVQNOHMLAKKLEDFEMLYA	420
Qy	1026	HGPDLCRESDDLHNAMANCEFALIGAVYLEGSLSEAKQLFGRLLFNDPDLREWLNYPLP	1085
Dd	421	HGPDLCRESDDLHNAMANCEFALIGAVYLEGSLSEAKQLFGRLLFNDPDLREWLNYPLP	480
Qy	1086	LQLOEPTDRLQILTSVYLQKLEFEELAGVFTYHRLLRAPFLRYGFFNLTLGHNR	1145
Dd	481	LQLOEPTDRLQILTSVYLQKLEFEELAGVFTYHRLLRAPFLRYGFFNLTLGHNR	540
Qy	1146	MEFLGDSIMOLVATEYELFHFPPDHGHGHTLLRSSLYNNRQAKVAEELGMEVATITDK	1205
Dd	541	MEFLGDSIMOLVATEYELFHFPPDHGHGHTLLRSSLYNNRQAKVAEELGMEVATITDK	600
Qy	1206	TKRPVGLRTKTLADLLESLFTALYTKTDLEYVTFMNVCFPRLKEFLINDQNDPKSQL	1265
Dd	601	TKRPVGLRTKTLADLLESLFTALYTKTDLEYVTFMNVCFPRLKEFLINDQNDPKSQL	660
Qy	1266	QOCCLTIRTEBKEKEDIPLYKTLQTVGSPHARTYAVVYKGERIGCGKPSIQQAEMGNA	1325
Dd	661	QOCCLTIRTEBKEKEDIPLYKTLQTVGSPHARTYAVVYKGERIGCGKPSIQQAEMGNA	720
Qy	1326	MDALEKINFPQMAHQKREIGRKTRQELKEMRERHOREPDETDIKK	1374
Dd	721	MDALEKINFPQMAHQKREIGRKTRQELKEMRERHOREPDETDIKK	769
RESULT 3			
09XYNS PRELIMINARY: PRT: 1327 AA.			
ID	09XYNS		
AC	09XYNS;		
DT	01-NOV-1999 (Tremblrel. 12, Created)		
DT	01-NOV-1999 (Tremblrel. 12, last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, last annotation update)		
DE	DROSHA PROTEIN.		
GN	DROSHA OR CG8730.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
NCBI_TaxId=7227;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RC	MEDLINE=20136006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amann-Rad A.C., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yiannelli M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,		
RA	Man K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,		
RA	Abell J.F., Aghayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,		
RA	Balcer R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Davis A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,		
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,		

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
39.5%; Score 2963; DB 5; Length 1327;	45.6%; Pred. No. 5,4e+202;	621;	213;	327;	200;	32;
Query	90	CGPRLPPCPRLRPPFPNMOKRHPRPVPCPPMPMPNPN-CN-NPPVYGAP-----P	137			
Db	3	QPPRLPPVQPA-----PPPPPPPEEDLSPPGVGPSNHYSSNESHQ	46			
Qy	138	GGGTFFPMKPPRSMRHPRRPVMPQOVNVQYVPGY-----SHHNPP	179			
Db	47	SKSLDYVPEPRPAVYASS---VPSTYDPRQDPRAVGYEGYALNEDQAKYGGGSEIYQY	103			
Qy	180	PSFNSFQNNPSSFLPSANNSSPHFRHLPPVPL---KPADESRSPERLKHYYDHRHD	235			
Db	104	PA-----SGSSFLYES-----YKPYDRXPAVYSSNTRPSERO-----RYTS	139			
Qy	236	HSHGGERIRSLDRERGRSPDRRQDSDRYRSDYDRGRTPSNHRSEYKRSRERERRNHR	295			
Db	140	NSSSGCYHNYP-----GYSSGR-----RYEQRRHQ-----EHRQIDQSRANRPHGHY	183			
Qy	296	DNRSPSLRSEYKKEKYSKRSRGSYGLSVVDEPAGCTPELPGELIKTWDMAAPLFTVHRS	355			
Db	184	AHRQAKGSOHG---YGSARN-----QVSDIYSPRGHNERRN	219			
Qy	356	PSREKKRA--RWEKKR---WSDNOSGDKDNTYSIKEKEPEDETMDCNKEEELLLK	409			
Db	220	ETLEKTRAKPKVETBRDLRQWCSNFC-----EKPEDYVKKMAALSPADAPV	267			
Qy	410	PWVICTHSENTYSSDPMDQYDSTIVVGTSHRLDLYKFEEDLSRQEKAKAARPPERP	469			
Db	268	ESWVYSSPAELIYETK---SENVVRAROROKTCTFDELLQRAKRVKELPPVVP	323			





```
|||||
Db 242 EDLESSESECESEDSTSCSSSDSEFVIAIEIKKKAKHAPRLDELWYNDPGQMDGP 301
OY 535 LCKCSAKARTGIRHSIYGEAIIKPCRPMTNNAGRLPRTTVPNNFLDRTVLEY 594
Db 302 LCKCSAKARTGIRHSIYGEAIIKPCRPMTNNAGRLPRTTVPNNFLDRTVLEY 361
OY 595 DDHEYIFEGFSMFANAPLT 613
Db 362 DDHEYIFEGFSMFANAPLT 380

RESULT 7
001327 PRELIMINARY; PRT; 1086 AA.
ID 001327
AC 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
GN F26E4.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
RM
RP SEQUENCE FROM N.A.
RA Lightning J.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; 281070; CAB03006.3; .
DR InterPro; IPR001159; DS_RBD.
DR Pfam; PF00035; dsrm. 1.
DR SMART; SM00358; dsrm. 1.
DR SMART; SM00353; RIBOC; 2.
DR PROSITE; PS00137; DS_RBD; 1.
DR PROSITE; PS00517; RNASE_3_1; 2.
DR PROSITE; PS0142; RNASE_3_2; 2.
SQ SEQUENCE 1086 AA; 12533 MW; 4A478120F88F8FB8 CRC64;

Query Match 19.3%; Score 1448; DB 5; Length 1086;
Best Local Similarity 31.2%; Pred. No. 2.7e-94;
Matches 364; Conservative 206; Mismatches 431; Indels 166; Gaps 31;

OY 277 RHRSESRERERHRRHRRSPSLERSYKKEYKASGRSYGSLVPEPAGCTPELP-G 335
Db 14 KKKRRARRKKYQAEYQBRKHEEMQOLGRRFQNPSTSSAPPTYKLTLPRESTALDFG 73
OY 336 EIKKNTDS-----WAPPLEIVNHRSPSRKKRAMEEKDRMSDQSGKDKNTYSIK 388
Db 74 DSPRLTEKDYETNYMIDPP--VVSFHSALIK-----SNRVYIK 110
OY 389 EKEPEETMPDKNEEEELK-----PWTIRCHS----- 418
Db 111 ABEAKYVMIKKSTTSTKIDFQKILETVYKTRRLQADVPYIILPHCSMKGRKTPKQK 170
OY 419 ---ENYSSDPMQVGDSTVVGTSRLDLVDFEELSGROEKAARPP--WEPPKTKL 473
Db 171 GGDEFTASDVSDSDSND-----QDEASTSPTPRQAPLEADKT 208
OY 474 DEDLESSESECESEDSTSCS-----SSDSEVFVYIAIKKKAKHAPRLH 519
Db 209 GR-----YKDEKQTCNRRNOQRAKRLRNPFKEKEQITLLKKIGIDRKKTHPNQIH 258
OY 520 DELWYNDPGQMDGPLCKCSAKARTGIRHSIYGEAIIKPCRPMTNNAGRLPHY--RIT 577
```

```

Db 259 PDIFENKGLGNEBPECRCPEPIKTGLKHYGAGEDKALDCK--KSNGENLHYTLRVT 316
OY 578 VSPPTNLTRPVIYIEDDHEYIFEGFSMFANAPL-----TNPLCYVIRNIDYTHFIE 633
Db 317 PLPSENOL--YRTHMALNGEFEFEFGSLTTHAPLPDCMRADICK---YSMDYQLEVE 371
OY 634 EMMP-ENFCVKGLEFLSLFELFRODILEYDWNLGKPLFEDSPCCPRHMFPRVRLPVG 692
Db 372 EFMFDECFDEDCDMLFEYTFHELFEMDLDELKPHIPSVYECSPHIIHMPRVQ--TKKD 430
OY 693 GREVLSMHQLLYLL--RCSKALVPEEELIAMLOMELEWQYAECKGMIVTNGTPSS 751
Db 431 LVQLMSSKTVLAFTSKSGSEIMSPEDVNRICDAQIDQFTRNSKHOSIVLTNTEKPESA 490
OY 752 VRIDQLRDEQFNPDPVITFPIIVHNGIRPADLSVAGSDPOYQKMKSVYKLRLHLNAPYK 811
Db 491 IRADWFERDEKKEV---YVHNAIRAQYTYTAISLPRIAFLEKTYLAKMKIQ--EKSSGVY 545
OY 812 QTDKOKLAOREEALOKIRQKNTMRREYVELSSQGEWKTGIRSDVCQHAMMLFVLNHR 871
Db 546 NKDEFK---TKNELEHLKRENRBSARNKLREPVAGFTENGKRDVAHVMTTLACHHR 602
OY 872 YHOCMLHDLKIGTFODRCLDLAMTHPSHILNFGMNDPHARNSLSCGIRQPKYG--D 929
Db 603 YNFSLDVFEERYEYKFNDRVIEIALMHSFSGHYGTPIDHVKMINTNGCYRR--KYGARD 661
OY 930 RKVYHNMRRKKGINTLINISRLGODDPPRSINNHRELEFGDAVEFLTSVHLTYLP 989
Db 662 KR--EKRRVAGIWSLRIKNGKTSGGE--TLHNERLEYLDADAVELISNHLTYFMTL 714
OY 990 SLEBGLATYRTAIVONHNLAMLAKELEDPFMYLVAGHGDLCRESDLRHAMAKCFEALIG 1049
Db 715 HHFEGGLATYRTALVQNRNLATLAKNCRIDEMQLQYSHGDLIVAGFEKHALANAEFVVA 774
OY 1050 AVYIEGSLERAKQLFGRLF--NDPDLREWLNTPLPRLQLEPNTROLIETSPVLOKIT 1108
Db 775 AIYLDGLIAPCDVIFSKAMYGHPVLKEKWDHINEHLKEDEPQGDRIISFPIITLSTPH 834
OY 1109 EFEELGIVTFHVLARATFRTVGGFHNHLLTGHNOEMERLEFGDAVEFLTSVHLTYLP 1168
Db 835 ALERRLGIOFNTRILAKATATRRNIPNDLTGHNOERLEFGDAVEFLTSVHLTYLP 894
OY 1169 HHGHLTLRSLVNNNTQAKVAEELGMOEYATINDTKRPV---GLRKTTLADLLESP 1225
Db 895 HHGHSMLTSLVSNQTOQVVDGDTGTFEVL-----KAPYTPRLKLDKADIVLEAF 949
OY 1226 AALYTDKDELYVATFMANVCFPRFLKEFILNQDNDNRKSQLOCCCLLR--TEGKEPDILY 1284
Db 950 GALYVDGIGHCRAFRIRYPCPRKHFIESEKYNDAKSHLOQCLAMRPPSSSEPDMPY 1009
OY 1285 KTLQTVGSHARTYTVAVVYKGERIGCGKPSIQOAEQMGAMALRYNFPQWAKRFT 1344
Db 1010 RVLGIEPTNNRIFKTAIVYIKGKRILASAEISYHAKELVAYELALA--NLEKSPSK--- 1064
OY 1345 GRKYROELKEMREHOREPDETD 1371
Db 1065 -----MKAKNNSWFOQNRRLLEDQTS 1086

RESULT 8
090908 PRELIMINARY; PRT; 860 AA.
ID 090908
AC 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
GN RIBONUCLEASE (FRAGMENT).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
```

RP SEQUENCE FROM N.A.  
 RC STRAIN-CH1489 HIM-8(E1489);  
 RX MEDLINE-20179701; PubMed-10713462;  
 RA Filippov V., Solovyev V., Filippova M., Gill S.S.;  
 RT "A novel type of RNase III family proteins in eukaryotes."  
 RL Gene 245:213-221(2000).  
 DR EMBL: AF160248; AAD45518.1; -  
 DR InterPro: IPR001159; DS\_RBD.  
 DR InterPro: IPR000999; RNase\_3.  
 DR Pfam: PF00035; dsrm; 1.  
 DR Pfam: PF00636; Ribonuclease\_3; 2.  
 DR SMART: SM00358; DSRM; 1.  
 DR SMART: SM00355; RIBOC; 2.  
 DR PROSITE: PS0137; DS\_RBD; 1.  
 DR PROSITE: PS00517; RNASE\_3.1; 2.  
 DR PROSITE: PS0142; RNASE\_3\_2; 2.  
 FT NON\_TER 1  
 SQ SEQUENCE 860 AA; 99414 MW; B339A3E7C9622A08 CRC64;

Query Match 18.9%; Score 1417; DB 5; Length 860;  
 Best Local Similarity 36.2%; Pred. No. 3,2e-92;  
 Matches 318; Conservative 173; Mismatches 334; Indels 54; Gaps 21;

QY 508 IRRKKAHPRLDELWYNDPGQMDGPKCSAKARRTGIRHSIYEGEEAIKPCRPMTNN 567  
 DB 21 IDRRKTHPNCIHDPDISFENEGKLGNEGECRCPEPIKTCGLHGUYAGEDKADICK--KSN 78  
 QY 568 AGRLEFH--RTVSPPTNLTDRPTVYEDDHEIYEGEGSMFNAAPL---TNPICLKYI 621  
 DB 79 GENHHTLVLPSPENQ--YRTNHAINGEEFEFGSLTHADPDMCTAPKIC-- 134  
 QY 622 RFNIDYTHIEEMMP-ENFCVKGLELSFLRDLILEYDMNLKPLFEDSPCCPRFH 680  
 DB 135 -YSMDYEQVLEERMPDECDPDCDMLFEYIFHEIFEMDFELRKPHIPSDVSECPMH 193  
 QY 681 EMRFVFLFDGGEVLSMAQIILYTL-RSKALVPEEETANMLQMLEMOKYAECKG 739  
 DB 194 IMPRFVQ-TRDDVLQVMSKTVLAYFTSKSSSEIMSPEDVNRCLDQIDQFTNTSKHKQ 252  
 QY 740 MIVNPTKSSVYRIDLDREOFNPVITPTIVHGGIRPAQSVAGDPQYQKLMKSYK 799  
 DB 253 SYLVNTFKSAIRADWFEDKEKEY---YVHNAIRAQTYTALSIPRIAELEKTLNK 308  
 QY 800 LRHLANSPVKQTDOKLAQREBALOKIKQKTMREYVELSSQGFMTGIRSDYQKH 859  
 DB 309 MIO-EKQSSGVYKNDK--TKNELEHLKREKRSANLKLRPVAGFETGKLPDYAAH 364  
 QY 860 AMMLPVLTHIRTHQCLMDKLGTFQDRCILQAMTHPSHNLGNMPPDHARNSLN 919  
 DB 365 VVMTILCHHIRYVFSIDVEEYEVYKFNDRVYELALMHSFKSHGTPRIDHVKNMTN 424  
 QY 920 CGTROPYUG--DRKVHNMHRKKGINFLIMSLGDDTPSRINNELEFLGAYVE 977  
 DB 425 CGIRK-KYGAEDKR--EKKRVAGIMSLFNMGTSGEP---LHNELEVLGAYVE 476  
 QY 978 FLTSVHLUYFSPLEEGGLATYRTAIVQNOHNLAMAKLELDPMLYAHPPDLCRESDLR 1037  
 DB 477 LYSVHNLUYFMTLHFEGLATYRTALVQNNLNTLAKNCIDEMLQYSHADLINAERK 536  
 QY 1038 HAMANCFEALIGAVYEGSLSEAKQFGRLLF-NDPDLRFVWLYNPLRLQLOEPNTDQ 1096  
 DB 537 HALANAEAVMAALYLDGLGAPCDVIFSKAMYGHQPLAKEMKHINHEHEKREDPQDRD 596  
 QY 1097 LIESPVLQKLEFEELAGIYFTFHVRLARAFILRTYGFNHLTGNQNRKEFLGDSIMOL 1156  
 DB 597 LSFPTPLTFPRLHLEERLGIQFNIRLLAKAFTRRNLPNDLTKGHQRLWEMLDGSDVLQ 656  
 QY 1157 VATEYLFTHRPDHEGLTLTSSLVNNRTQAKVAEELAGQEAITYTNDKTKRPV---GLR 1213  
 DB 657 IVSDFLVRPRYHNEGMSLRTSLVSNQTOAVVCDLGLTFEYI-----KAYYKPELK 711  
 QY 1214 TKTLADLSEFTALYTDKDLVEYHTEMVCFPRLEKEFLINDQMDPKSQLOQCCLTLR 1273  
 DB 1214 TKTLADLSEFTALYTDKDLVEYHTEMVCFPRLEKEFLINDQMDPKSQLOQCCLTLR 1273

DB 712 LKDKADLVEAFIGALYVDRGIEHCRAFLIRVCPRLKHFISEKWNDAKSHLQOQCLAMR 771  
 QY 1274 -TEGKPPDPLTKTLOTGSPSHARTTYAVYFKRGICGKGPSTQOAEQMAADALEKY 1332  
 DB 772 DSSSEPPMPPEPRVLGICIEPTNNRIFKTAIVYKKGRLASAEASNVRKALVAVELALA-- 829  
 QY 1333 NFPQAHQKRFIGRRYQOLKEMRWEREHREPERDETD 1371  
 DB 830 NLEMSFSK-----MKAKNSWFOQNMRRRLJEDTSD 860

## RESULT 9

Q9CTG2 PRELIMINARY; PRT; 134 AA.

AC 09CTG2;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE 1110013A17RIK PROTEIN (FRAGMENT).  
 GN 1110013A17RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J; TISSUE-EMBRYO;  
 RX MEDLINE-21085660; PubMed-11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
 RA Felschmann W., Gaasterland T., Gissi C., King B., Kochava H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staabli F., Suzuki K., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guslinich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayshtshak Y.  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK003651; BAB22917.1; -  
 DR MGD: MGI:1915895; 1110013A17RIK.  
 DR InterPro: IPR001159; DS\_RBD.  
 DR Pfam: PF00035; dsrm; 1.  
 DR SMART: SM00358; DSRM; 1.  
 DR PROSITE: PS0137; DS\_RBD; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 134 AA; 15883 MW; 72AA673BFDCA1C6 CRC64;

Query Match 9.5%; Score 711; DB 11; Length 134;  
 Best Local Similarity 97.8%; Pred. No. 4.1e-43;  
 Matches 131; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1241 MNVCFPRLEKEFLINDQMDPKSQLOQCCLTLRTGKRPDILYKTLQTVGSPSHARTTYV 1300  
 DB 1 MNVCFPRLEKEFLINDQMDPKSQLOQCCLTLRTGKRPDILYKTLQTVGSPSHARTTYV 60  
 QY 1301 AVYFKERIGCGKGPSTQOAEQMAADALEKYNFPQAHQKRFIGRRYQOLKEMRWERE 1360  
 DB 61 AVYFKERIGCGKGPSTQOAEQMAADALEKYNFPQAHQKRFIBRYKROELKEMRWERE 120  
 QY 1361 HOEREPDEDEDIKK 1374  
 DB 121 HOEREPDEDEDIKK 134

```

RESULT 10
OSVAV4 PRELIMINARY: PRT: 1151 AA.
ID OSVAV4
AC OSVAV4:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CG5514 PROTEIN.
GN CG5514.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharya D., Bolshakov S.,
RA Borovda D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hsolin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinslock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003763; AAF56763.1; -.
DR Flybase; FBgn0039560; CG5514.
DR InterPro; IPR002965; P_richextensn.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 1151 AA; 125450 MW; AF8330378658C1D1 CRC64;
Query Match 4.5%; Score 340; DB 5; Length 1151;
Best Local Similarity 22.5%; Pred. No. 2,1e-15;
Matches 173; Conservative 83; Mismatches 280; Indels 234; Gaps 35;
OY 28 RSADSPFRQNLRLHPQOPVQYQYEPSPASSTTFSSNPAVNF-----PPRPDVFPP 82
DB 362 RASPFVEPP-----PPAPGVESPPFPQPPASPRFPDPHPHTIEPPPPAPPLVAPP 415
OY 83 PPMPSAGQPLRPPC--IRPPRNQMKRHPFVPPCFPPPPPPMKCPN--NPPVPGAP-- 136

```

```

DB 416 PPAPPTIKPPPPAPPTVEPPP-----PPAPPTVEPPPPPPAPPTKVEPPPPAPAE 469
OY 137 --PGQGTFFPPMMPPSPMHP-----PPVMPQOVNTQYPP-----GY----- 172
DB 470 VEPPPAPPTLEPPPPPPAPPKVELPPPPAPPAEALITPRKAGSGFTTELAATPKES 539
OY 173 ---SHNPF-----PPSFNSFQNNP-----SFLPSANSSSPH 203
DB 530 ETROKPHNVMTDYKEDQBPATTEDEKQKPLEGLDAPKDAVSTASSTESTSTSSOS-- 587
OY 204 FRHLPPPLKAPBESRPERLKHVDHRHDSHGEGHRLDR--REKGRSPDDRROD 262
DB 588 -----KSHSKSDKEKEKDRHRHRSDDKH-----RRSTDHROKSRDKSHSKS 633
OY 263 SRRYSDYDRG--PPSRHSYERERERERHRHRDRNRSPSLR-----SYRK 309
DB 634 SSSSSSKSSSSSSSKKSSSSSKNDKSSSSSSSRSSSSSSSGTSSSRSSSSSHK 693
OY 310 -----EYKRSGRSYGLSVPEPAGCPPELPGETIKMTDSWAPPLETVNHRSPREKRRAR 364
DB 694 HKSSSSSSSRSEDRKG-----KEKDKEREKDSQRS-----HHSSSSSSSSSR 736
OY 365 WEEKDKWD-----NQSCKDN-----YTSYE-----KEPETMDKNE 401
DB 737 KDHRGRDRNRKNSNTSGAENKAIHDDHSESEKRYQRGRSDSNDGKRPSSGPAKNS 796
OY 402 EEEDE--LKPVWIRCTHSENYSSDPMDQVDS---TVGTSRLDYDKFEELGSR 455
DB 797 QPESAAITMSDPAVEANANTNGNSNGTACADNVSGIYVTDIIOQSTSSVELTAAS 856
OY 456 QEKARAPPEPPPKTKLDELESSSECESEDSTCSSSD-----SEFPDVAIEIK 509
DB 857 QSHRAASSKHEPEDIDGKEADNQPEKTELEARQDE--CASQNEVPVPEPQTLADSVPLP 915
OY 510 RKRAHRLDELMTYNDPQGMNDGPLCKSARKRGIRISITPGEATIKPCPMPNNMG 569
DB 916 AKES--MTLEDE-----KETNVEBN--KSEESKP----- 943
OY 570 RLPHYRTVSPPTNLTDRPTVLEYDDHEYIFGFSMFAHAPLTNPLC--KVIRPNIDY 627
DB 944 -----ENPPECVDEPAGQVDPEP-----PQDDAKPATPVPIISMEQDEFSADF 988
OY 628 TIFTEEMPEPNCVKGLEFLFLFRDILLEYDW-----NKGLEFLFD 671
DB 989 VTHF--EENTDE-----FTRRLQNLINOLEDKRNLNLSIED 1023
RESULT 11
ID Q16630 PRELIMINARY: PRT: 551 AA.
AC Q16630;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HBBRII-4 mRNA.
GN HBBRII-4 OR HBBRII-7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fleischhauer K.L.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBD databases.
DR EMBL; X67337; CAA47752.1; -.
DR EMBL; X67336; CAA47751.1; -.
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR POSITE; PS50102; RRM; 1.
SQ SEQUENCE 551 AA; 59208 MW; 407CB941BA56A966 CRC64;
Query Match 4.2%; Score 316; DB 4; Length 551;

```









```
Db 889 TPRKASHSKAKEHOEAKPAKDEKVKKDCSKDIKSEKPAK--DEKAKKPEKNKLLDSKGEK 947
Qy 545 -----TGIRHSITVGEELIKPC--RPMNNNGRLFH-----YKITYSP 580
Db 948 RKRTEEEKVDKDFESSKISKVEGETELVKPSPKRMKGDEKLERPEKDKIASSTTP 1007
Qy 581 PTNFLDRPTVLEYDDHEIFEGFSMAFAPLNIPLCVIFNFIDYTHFTEEMMPENF 640
Db 1008 AKKIKINRETGKKIGNAE-----NASTTKPSEK-----LESISKIKÖKGVK 1052
Qy 641 C---VKGLEFLSFLFRDILLEYDMNLKGLFEDSPCCRFHFMPFRFLPDGKEVL 697
Db 1053 AKRKVAGSESSSTLY-----DYSTSGTSGSPVR 1082
Qy 698 SMHOLLILYLRCSKAL-----VPEERIANMLQ-----WEELEOKAECK----- 738
Db 1083 KSEKTDYTKRYVYIKTMEEYNNNDTAPAEDEVITMIOVPSKWDKDDFEESEEDVKTTOPIQ 1142
Qy 739 -----GMTVNFPTKPSRYRIDLDREOPNDVITFPRIIVHGIRPALSYAGDPOYOKL 793
Db 1143 SVGKPSIITIKNTYTKPSAT--AKYTEKESQPE-----KLOKL 1178
Qy 794 WK--STVKLRHL-----ANSPKYQTDKQKLAOREALQIRKQNTMRREYVE--LS 844
Db 1179 PKASHHELMQHEIRSKSGASSEKGRAKDREHSGSEKNDPKRKSGAQDPKESTVDRLSE 1238
Qy 845 QGFWKT 850
Db 1239 QGFHKT 1244
```

## RESULT 15

Q9SBM1

ID Q9SBM1 PRELIMINARY: PRT: 409 AA.

```
AC Q9SBM1; 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP PRECURSOR.
GN HRGP GENE.
OS Volvox carterl f. nagariensis.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID-3068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK10;
RX MEDLINE-20044763; PubMed-10574980;
RA Ender F., Hallmann A., Amon P., Sumper M.;
RT "Response to the sexual pheromone and wounding in the green alga
RT Volvox: induction of an extracellular glycoprotein consisting almost
RT exclusively of hydroxyproline."
RL J. Biol. Chem. 274:35023-35028(1999).
DR EMBL; AJ242540; CAB62280.1; -.
DR InterPro: IPR003882; Pistil_extensin.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR01218; PSTLEXTENSIN.
KW signal.
FT SIGNAL. 1 17
FT CHAIN 18 409 HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP.
SQ SEQUENCE 409 AA: 41547 MW: CD0749C6AF02BD74 CRC64;
```

## Query Match

Best Local Similarity 3.8%; Score 285; DB 10; Length 409;

Matches 76; Conservative 11; Mismatches 79; Indels 32; Gaps 6;

```
Qy 31 APSPFRQNLRLHPQRPVQYQYEPSSA--PSTTFSNSPAPNPLPRPRDF-----VPFRPP 84
Db 19 APARKRP-----PRSPVALVETPAAPRGSPRGTPRGVPRPTPSGPHRPPRPP 72
Qy 85 MPSSAGPLRPPCIRPPFNHQRHPRFVVRPCFPRPMPRPPCRRNRPVVGARPGGTFFP 144
```

```
Db 73 PPPPPQPLPPSPSPPPPP-----PPVPPPPPPPPPPPPSPPPPPPP----- 119
Qy 145 MMPSPSMHPPPPPVVMQOVNYYQYPPGISHHNFPPPSFNSFQNNPSSFLSANNSSPHF 204
Db 120 --PPSPSPPPPPPPPPPPPPPPPPPP-----PPSPSPSPSPSPSPSPSPSPSP 172
Qy 205 RHLPPYLPKAPSERRSP 222
Db 173 PPSPPSPSPSPPPPPPP 190
```

Search completed: October 9, 2002, 18:25:49  
Job time : 100 secs

